

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 11:08:54 ; Search time 3842.15 Seconds
(without alignments)
1764.726 Million cell updates/sec

Title: US-09-394-745-6603
Perfect score: 411
Sequence: 1 agcaaaagcatagagatcca.....aggagaagaggaagggaccg 411

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*

28: em_un:*
 29: em_vi:*
 30: em_htgo_hum:*
 31: em_htgo_inv:*
 32: em_htgo_rod:*
 33: em_htg_hum:*
 34: em_htg_inv:*
 35: em_htg_rod:*
 36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	% Query						Description
	No.	Score	Match	Length	DB	ID	
c	1	45	10.9	2217	3	SAHNRNPH	X54670 S. american
c	2	42.4	10.3	27570	2	AC069197	AC069197 Homo sapi
c	3	42.2	10.3	76734	2	AC023218	AC023218 Homo sapi
	4	41.4	10.1	45313	1	SCD95A	AL357432 Streptomy
	5	41.2	10.0	60609	2	AC023519	AC023519 Homo sapi
c	6	40.8	9.9	102242	2	AP004043	AP004043 Oryza sat
c	7	40.8	9.9	125118	2	AP003865	AP003865 Oryza sat
c	8	40.2	9.8	3338	9	AF233344	AF233344 Homo sapi
c	9	40.2	9.8	179004	2	AC009988	AC009988 Homo sapi
	10	39.8	9.7	11096	1	AF275943	AF275943 Streptomy
c	11	39.8	9.7	56494	2	AC083782	AC083782 Homo sapi
c	12	39.6	9.6	36734	1	SC6G10	AL049497 Streptomy
c	13	39.6	9.6	78220	2	AC023212	AC023212 Homo sapi
c	14	39.2	9.5	54450	2	AC080179	AC080179 Homo sapi
c	15	39	9.5	11070	1	AE005747	AE005747 Caulobact
	16	39	9.5	19791	1	SPFKBAD	Y10438 Streptomyce
c	17	39	9.5	110000	2	HSS171M_2	Continuation (3 of
c	18	39	9.5	340000	9	HS21C102	AL163302 Homo sapi
c	19	38.8	9.4	63121	2	AC036107	AC036107 Homo sapi
	20	38.6	9.4	198218	2	AC037447	AC037447 Homo sapi
	21	38.4	9.3	158063	9	AP001046	AP001046 Homo sapi
c	22	38.4	9.3	179556	2	AC018734	AC018734 Homo sapi
c	23	38.4	9.3	219278	2	AC016018	AC016018 Mus muscu
	24	38.4	9.3	340000	9	AP001751	AP001751 Homo sapi
	25	38.2	9.3	2378	1	AF184600	AF184600 Streptomy
	26	38.2	9.3	2507	10	RATVGFA	M74223 Rat VGF mRN
	27	38.2	9.3	68902	2	AC091585	AC091585 Homo sapi
c	28	38	9.2	67477	2	AC055848	AC055848 Homo sapi
	29	38	9.2	87076	9	AC005918	AC005918 Homo sapi
c	30	37.8	9.2	220469	2	AC074307	AC074307 Mus muscu
	31	37.6	9.1	83393	2	AC087331_3	Continuation (4 of
	32	37.4	9.1	28732	1	AF082100	AF082100 Streptomy
	33	37.2	9.1	141279	2	AC023824	AC023824 Homo sapi
c	34	37.2	9.1	298166	2	AC087563	AC087563 Homo sapi
c	35	37	9.0	10986	1	AE003962	AE003962 Xylella f
c	36	37	9.0	15776	1	AE007069	AE007069 Mycobacte
	37	37	9.0	19830	1	SC3F9	AL023862 Streptomy
c	38	37	9.0	37586	1	MTCY270	Z95388 Mycobacteri

c	39	37	9.0	37586	6	AX191745	AX191745 Sequence
c	40	37	9.0	76336	2	AC023249	AC023249 Homo sapi
c	41	37	9.0	223940	2	AC087567	AC087567 Mus muscu
	42	36.8	9.0	9963	1	AE005878	AE005878 Caulobact
	43	36.8	9.0	63449	2	AC026976	AC026976 Homo sapi
	44	36.8	9.0	82024	2	AC023210	AC023210 Homo sapi
c	45	36.8	9.0	163275	2	AC007623	AC007623 Homo sapi

ALIGNMENTS

```

RESULT 1
SAHNRNPH/c
LOCUS SAHNRNPH 2217 bp mRNA INV 04-MAR-1991
DEFINITION S. americana hnRNP mRNA for protein homologous to A1, A2/B1
            proteins of mammalian hnRNP.
ACCESSION X54670
VERSION X54670.1 GI:10106
KEYWORDS hnRNP A1 protein; hnRNP A2/B1 protein; hnRNP protein.
SOURCE American grasshopper.
  ORGANISM Schistocerca americana
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;
            Acridomorpha; Acridoidea; Acrididae; Schistocerca.
REFERENCE 1 (bases 1 to 2217)
  AUTHORS Ball,E.E.
  TITLE Direct Submission
  JOURNAL Submitted (29-AUG-1990) Ball E.E., Australian National University,
            Molecular Neurobiology Group RSBS, P O Box 475, Canberra City
            A.C.T. 2601, Australia
REFERENCE 2 (bases 1 to 2217)
  AUTHORS Ball,E.E., Rehm,E.J. and Goodman,C.S.
  TITLE Cloning of a grasshopper cDNA coding for a protein homologous to
            the A1, A2/B1 proteins of mammalian hnRNP
  JOURNAL Nucleic Acids Res. 19 (2), 397 (1991)
  MEDLINE 91195067
FEATURES             Location/Qualifiers
     source            1. .2217
                       /organism="Schistocerca americana"
                       /db_xref="taxon:7009"
                       /dev_stage="35-50% embryo"
                       /clone_lib="gt11"
     mRNA              <94. .2217
                       /gene="hnRNP"
                       /note="putative"
     gene              94. .2217
                       /gene="hnRNP"
     CDS               94. .1122
                       /gene="hnRNP"
                       /note="putative"
                       /codon_start=1
                       /product="mammalian A1, A2 /B1 hnRNP homologue"
                       /protein_id="CAA38481.1"
                       /db_xref="GI:10107"
                       /db_xref="SWISS-PROT:P21522"
                       /translation="MVNMDRENDHGEPEHVRKLFIGGLDYRTTDESLKQHFQWGEIV

```

DVVVMKDPKTKRSRGFGFITYSRAH MVDDAQNARPHKVDGRVVEPKRAVPRTEIGRPE
AGATVKKLFVGGIKEEMEENDLRDYFKQYGT VVSAAIVVDKETRKKRGFAFVEFDDYD
PVDKICLSRNHQIRGKHIDVKKALPKGDAPGGRGGGGGGGGVGGGAGGGWGGGRGDWGG
SAGGGGGGGWGGADPWENGRGGGGDRWGGGGGGMGGGDRWGGGGGGMGGGDRYGGGGGR
SGGWSNDGYNSGPQSDGFGGGYKQSYGGGAVRGSSGYGGSRSAPYSDRGSRGGGGGGY
GSGGGRRY"

mat_peptide 94. .1119
/gene="hnRNP"
/note="putative"
/product="mammalian A1, A2 /B1 hnRNP homologue"

protein_bind 118. .375
/gene="hnRNP"
/note="containing RNP CS#1"
/bound_moiety="RNA"

misc_feature 265. .288
/gene="hnRNP"
/note="RNP CS#1"

protein_bind 391. .648
/gene="hnRNP"
/note="containing RNP CS#2"
/bound_moiety="RNA"

misc_feature 538. .561
/gene="hnRNP"
/note="RNP CS#2"

BASE COUNT 592 a 354 c 629 g 642 t
ORIGIN

Query Match 10.9%; Score 45; DB 3; Length 2217;
Best Local Similarity 51.8%; Pred. No. 0.16;
Matches 102; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

Qy 73 cctagcgataccaatacacccatcccaacactccaaaccaaccaacacttcaaccaaacc 132
|| | || ||| || |||| |||| | || |||| ||| | || | |||

Db 891 CCCTCCCATTCCACCACCCCCACCCACCTATCTCCACCACCCATACCGCCACCACCACC 832

Qy 133 accacaacaatgccttcagtaacccaggcccgctctcatgtggcgtagcgctcgcccgcggc 192
|| || | | | | ||| |||| |||| ||| ||| | ||| ||

Db 831 TCCCCATCGGTACCGCCACCACCACGGCCATTCTCCCATGGATCAGCACCTCCCCAGCC 772

Qy 193 atggtagagccccaccccttcgctcgcaatcccatcaccatgacccctcacgcctggcgcc 252
| || | || ||| || || || |||| | || | || |

Db 771 ACCACCTCCACCGCCACCAGCACTTCCACCCAGTCACCACGTCCTCCGCCCCAGCCACC 712

Qy 253 gccgccgacctctccaa 269
|| || | | ||||

Db 711 ACCTGCGCCACCACCAA 695

RESULT 2
AC069197/c
LOCUS AC069197 27570 bp DNA HTG 17-FEB-2001
DEFINITION Homo sapiens chromosome 15 clone RP11-505I24 map 15, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION AC069197
VERSION AC069197.3 GI:12958076
KEYWORDS HTG; HTGS_PHASE0.

SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 27570)

AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE Homo sapiens chromosome 15, clone RP11-505I24

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 27570)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (21-MAY-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT On Feb 17, 2001 this sequence version replaced gi:11120926.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L9315
Center clone name: 505_I_24

* NOTE: This record contains 34 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

```

*      1      719: contig of 719 bp in length
*      720 819: gap of      100 bp
*      820      1529: contig of 710 bp in length
*     1530 1629: gap of      100 bp
*     1630      2340: contig of 711 bp in length
*     2341 2440: gap of      100 bp
*     2441      3132: contig of 692 bp in length
*     3133 3232: gap of      100 bp
*     3233      3971: contig of 739 bp in length
*     3972 4071: gap of      100 bp
*     4072      4818: contig of 747 bp in length
*     4819 4918: gap of      100 bp
*     4919      5620: contig of 702 bp in length
*     5621 5720: gap of      100 bp
*     5721      6419: contig of 699 bp in length
*     6420 6519: gap of      100 bp
*     6520      7235: contig of 716 bp in length
*     7236 7335: gap of      100 bp
*     7336      8075: contig of 740 bp in length
*     8076 8175: gap of      100 bp
*     8176      8878: contig of 703 bp in length
*     8879 8978: gap of      100 bp
*     8979      9704: contig of 726 bp in length
*     9705 9804: gap of      100 bp
*     9805     10508: contig of 704 bp in length
*    10509 10608: gap of      100 bp
*    10609     11332: contig of 724 bp in length
*    11333 11432: gap of      100 bp
*    11433     12149: contig of 717 bp in length
*    12150 12249: gap of      100 bp
*    12250     12963: contig of 714 bp in length
*    12964 13063: gap of      100 bp
*    13064     13791: contig of 728 bp in length
*    13792 13891: gap of      100 bp
*    13892     14564: contig of 673 bp in length
*    14565 14664: gap of      100 bp
*    14665     15357: contig of 693 bp in length
*    15358 15457: gap of      100 bp
*    15458     16191: contig of 734 bp in length
*    16192 16291: gap of      100 bp
*    16292     16999: contig of 708 bp in length
*    17000 17099: gap of      100 bp
*    17100     17822: contig of 723 bp in length
*    17823 17922: gap of      100 bp
*    17923     18639: contig of 717 bp in length
*    18640 18739: gap of      100 bp
*    18740     19458: contig of 719 bp in length
*    19459 19558: gap of      100 bp
*    19559     20275: contig of 717 bp in length
*    20276 20375: gap of      100 bp
*    20376     21104: contig of 729 bp in length
*    21105 21204: gap of      100 bp
*    21205     21938: contig of 734 bp in length
*    21939 22038: gap of      100 bp
*    22039     22756: contig of 718 bp in length
*    22757 22856: gap of      100 bp
*    22857     23573: contig of 717 bp in length

```

```

FEATURES                                Location/Qualifiers
    source                               1. .27570
                                         /organism="Homo sapiens"
                                         /db_xref="taxon:9606"
                                         /chromosome="15"
                                         /map="15"
                                         /clone="RP11-505I24"
                                         /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT      7225 a    4533 c    5245 g    7049 t    3518 others
ORIGIN

```

Query Match 10.3%; Score 42.4; DB 2; Length 27570;
Best Local Similarity 50.5%; Pred. No. 0.72;
Matches 97; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

```

Qy      76 agcgataccaatacaccccatcccaacactccaaaccaaccaacttcaaccaaaccacc 135
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 25940 ACCAACACACAAACACACACAACACAACACAAAACCAACCAACCACCAACAAAACCCAC 25881

Qy     136 acaacaatgccttcagtaaccagggcccggtctcatgtggcgtagcgctcgcccgcgggcatg 195
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 25880 AAACCACCACCCACCCACCCACCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 25821

Qy     196 gtagagccccacccttcgctcgcaatcccatcaccatgacccctcagcctggcgcgcc 255
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 25820 CCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCNCCNCCCCCCCCCCCC 25761

Qy     256 gccgacctctcc 267
      | | | | | | | |
Db 25760 CCCCCCCCCCCC 25749

```

```

RESULT      3
AC023218/c
LOCUS       AC023218       76734 bp       DNA             HTG             13-JUL-2000
DEFINITION  Homo sapiens chromosome 14 clone RP11-105M4 map 14, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC023218
VERSION     AC023218.2   GI:9165238
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
            ORGANISM      Homo sapiens
                        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                        Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1      (bases 1 to 76734)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.

```

TITLE Homo sapiens chromosome 14, clone RP11-105M4
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 76734)
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
 Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
 Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
 DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
 Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
 Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
 Landers,T., Lehoczký,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
 Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
 McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
 Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
 Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
 Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
 Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
 Zimmer,A. and Zody,M.

TITLE Direct Submission
 JOURNAL Submitted (09-FEB-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 COMMENT On Jul 13, 2000 this sequence version replaced gi:6957752.

All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L6632
 Center clone name: 105_M_4

* NOTE: This record contains 81 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

* 1 871: contig of 871 bp in length
 * 872 971: gap of 100 bp
 * 972 1838: contig of 867 bp in length
 * 1839 1938: gap of 100 bp
 * 1939 2804: contig of 866 bp in length
 * 2805 2904: gap of 100 bp
 * 2905 3736: contig of 832 bp in length
 * 3737 3836: gap of 100 bp
 * 3837 4667: contig of 831 bp in length
 * 4668 4767: gap of 100 bp


```

*      4768      5627: contig of 860 bp in length
*      5628 5727: gap of      100 bp
*      5728      6551: contig of 824 bp in length
*      6552 6651: gap of      100 bp
*      6652      7527: contig of 876 bp in length
*      7528 7627: gap of      100 bp
*      7628      8459: contig of 832 bp in length
*      8460 8559: gap of      100 bp
*      8560      9444: contig of 885 bp in length
*      9445 9544: gap of      100 bp
*      9545     10368: contig of 824 bp in length
*     10369 10468: gap of      100 bp
*     10469     11325: contig of 857 bp in length
*     11326 11425: gap of      100 bp
*     11426     12274: contig of 849 bp in length
*     12275 12374: gap of      100 bp
*     12375     13223: contig of 849 bp in length
*     13224 13323: gap of      100 bp
*     13324     14163: contig of 840 bp in length
*     14164 14263: gap of      100 bp
*     14264     15057: contig of 794 bp in length
*     15058 15157: gap of      100 bp
*     15158     16018: contig of 861 bp in length
*     16019 16118: gap of      100 bp
*     16119     16964: contig of 846 bp in length
*     16965 17064: gap of      100 bp
*     17065     17904: contig of 840 bp in length
*     17905 18004: gap of      100 bp
*     18005     18858: contig of 854 bp in length
*     18859 18958: gap of      100 bp
*     18959     19781: contig of 823 bp in length
*     19782 19881: gap of      100 bp
*     19882     20729: contig of 848 bp in length
*     20730 20829: gap of      100 bp
*     20830     21737: contig of 908 bp in length
*     21738 21837: gap of      100 bp
*     21838     22684: contig of 847 bp in length
*     22685 22784: gap of      100 bp
*     22785     23640: contig of 856 bp in length
*     23641 23740: gap of      100 bp
*     23741     24566: contig of 826 bp in length
*     24567 24666: gap of      100 bp
*     24667     25519: contig of 853 bp in length
*     25520 25619: gap of      100 bp
*     25620     26478: contig of 859 bp in length
*     26479 26578: gap of      100 bp
*     26579     27434: contig of 856 bp in length
*     27435 27534: gap of      100 bp
*     27535     28406: contig of 872 bp in length
*     28407 28506: gap of      100 bp
*     28507     29351: contig of 845 bp in length
*     29352 29451: gap of      100 bp
*     29452     30292: contig of 841 bp in length
*     30293 30392: gap of      100 bp
*     30393     31253: contig of 861 bp in length
*     31254 31353: gap of      100 bp
*     31354     32178: contig of 825 bp in length

```

* 32179 32278: gap of 100 bp
* 32279 33145: contig of 867 bp in length
* 33146 33245: gap of 100 bp
* 33246 34087: contig of 842 bp in length
* 34088 34187: gap of 100 bp
* 34188 35016: contig of 829 bp in length
* 35017 35116: gap of 100 bp
* 35117 35991: contig of 875 bp in length
* 35992 36091: gap of 100 bp
* 36092 36956: contig of 865 bp in length
* 36957 37056: gap of 100 bp
* 37057 37883: contig of 827 bp in length
* 37884 37983: gap of 100 bp
* 37984 38833: contig of 850 bp in length
* 38834 38933: gap of 100 bp
* 38934 39774: contig of 841 bp in length
* 39775 39874: gap of 100 bp
* 39875 40726: contig of 852 bp in length
* 40727 40826: gap of 100 bp
* 40827 41689: contig of 863 bp in length
* 41690 41789: gap of 100 bp
* 41790 42592: contig of 803 bp in length
* 42593 42692: gap of 100 bp
* 42693 43553: contig of 861 bp in length
* 43554 43653: gap of 100 bp
* 43654 44504: contig of 851 bp in length
* 44505 44604: gap of 100 bp
* 44605 45448: contig of 844 bp in length
* 45449 45548: gap of 100 bp
* 45549 46399: contig of 851 bp in length
* 46400 46499: gap of 100 bp
* 46500 47373: contig of 874 bp in length
* 47374 47473: gap of 100 bp
* 47474 48346: contig of 873 bp in length
* 48347 48446: gap of 100 bp
* 48447 49294: contig of 848 bp in length
* 49295 49394: gap of 100 bp
* 49395 50222: contig of 828 bp in length
* 50223 50322: gap of 100 bp
* 50323 51144: contig of 822 bp in length
* 51145 51244: gap of 100 bp
* 51245 52078: contig of 834 bp in length
* 52079 52178: gap of 100 bp
* 52179 53031: contig of 853 bp in length
* 53032 53131: gap of 100 bp
* 53132 53970: contig of 839 bp in length
* 53971 54070: gap of 100 bp
* 54071 54911: contig of 841 bp in length
* 54912 55011: gap of 100 bp
* 55012 55873: contig of 862 bp in length
* 55874 55973: gap of 100 bp
* 55974 56797: contig of 824 bp in length
* 56798 56897: gap of 100 bp
* 56898 57740: contig of 843 bp in length
* 57741 57840: gap of 100 bp
* 57841 58694: contig of 854 bp in length
* 58695 58794: gap of 100 bp

* 58795 59619: contig of 825 bp in length
 * 59620 59719: gap of 100 bp
 * 59720 60566: contig of 847 bp in length
 * 60567 60666: gap of 100 bp
 * 60667 61514: contig of 848 bp in length
 * 61515 61614: gap of 100 bp
 * 61615 62461: contig of 847 bp in length
 * 62462 62561: gap of 100 bp
 * 62562 63399: contig of 838 bp in length
 * 63400 63499: gap of 100 bp
 * 63500 64332: contig of 833 bp in length
 * 64333 64432: gap of 100 bp
 * 64433 65278: contig of 846 bp in length
 * 65279 65378: gap of 100 bp
 * 65379 66237: contig of 859 bp in length
 * 66238 66337: gap of 100 bp
 * 66338 67174: contig of 837 bp in length

Query Match 10.3%; Score 42.2; DB 2; Length 76734;
 Best Local Similarity 46.8%; Pred. No. 0.79;
 Matches 125; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

Qy 1 agcaaaagcatagagatccatcttctctgctcaatcaattacacaacaagagcattctag 60
 | | | | | | | | | | | | | | | | | | | | | |
 Db 21681 ACCCATAACATCCCCACCACANATAACAACCTAATAACCCACTCCCCTCCACCCTCTCAC 21622

Qy 61 atttgagttcatccttagcgataccaatacacccatcccaacactccaaaccaaccaacac 120
 | | | | | | | | | | | | | | | | | | | | | |
 Db 21621 CCCTCAAAACACCCACCCCAAACACACCCATCACACCCAACACTACCAATCTACCAACCA 21562

Qy 121 ttcaaccaaaccaccacacaacaatgccttcagtaaccaggcccggtctcatgtggcgtagc 180
 | | | | | | | | | | | | | | | | | | | | | |
 Db 21561 TCGAAACCCACCCCCAANACCACNCCACAACACCCNCCACACCAACTAATCTACCTAATC 21502

Qy 181 gtcgcccgcggcatggtagagccccacccttcgctcgcaatcccatcaccatgaccct 240
 | | | | | | | | | | | | | | | | | | | | | |
 Db 21501 CCTCTCCCCATCACTCACTACCACACACCCACCCTTCTCACCCCCCCTCCCACACCCCC 21442

Qy 241 cacgcctggcgcgccgacgtctcc 267
 | | | | | | | | | | | | | | | | | | | | | |
 Db 21441 CCCCCCATCCCCACCCACACACCCCC 21415

RESULT 4
 SCD95A
 LOCUS SCD95A 45313 bp DNA BCT 31-MAY-2000
 DEFINITION Streptomyces coelicolor cosmid D95A.
 ACCESSION AL357432
 VERSION AL357432.1 GI:8248766
 KEYWORDS acetyltransferase; ATP/GTP binding protein; carboxylesterase;
 chaperonin 2; cold shock protein; deacetylase; DNA-binding protein;
 helicase; histidine autokinase; homeostasis protein; integral
 membrane protein; metalloproteinase; mutT-like protein;
 NADP-dependent alcohol dehydrogenase; oxidoreductase; phosphatase;
 possible trehalose-phosphate synthase; reductase; regulatory
 protein; response regulatory protein; secreted protein; secreted
 solute-binding protein; sugar kinase; threonine synthase;

transcriptional regulator.

SOURCE Streptomyces coelicolor A3(2).

ORGANISM Streptomyces coelicolor A3(2)
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE 1 (bases 1 to 45313)

AUTHORS Redenbach,M., Kieser,H.M., Denapaite,D., Eichner,A., Cullum,J.,
Kinashi,H. and Hopwood,D.A.

TITLE A set of ordered cosmids and a detailed genetic and physical map
for the 8 Mb Streptomyces coelicolor A3(2) chromosome

JOURNAL Mol. Microbiol. 21 (1), 77-96 (1996)

MEDLINE 97000351

REFERENCE 2 (bases 1 to 45313)

AUTHORS Seeger,K.J. and Harris,D.

JOURNAL Unpublished

REFERENCE 3 (bases 1 to 45313)

AUTHORS Cerdeno,A.M., Parkhill,J., Barrell,B.G. and Rajandream,M.A.

TITLE Direct Submission

JOURNAL Submitted (31-MAY-2000) Streptomyces coelicolor sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
Colney, Norwich, Norfolk NR4 7UH, UK

COMMENT Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded
by the BBSRC and Beowulf Genomics
Details of S. coelicolor sequencing at the Sanger Centre are
available on the World Wide Web.
(URL; http://www.sanger.ac.uk/Projects/S_coelicolor/)
CDS are numbered using the following system eg SC7B7.01c. SC (S.
coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous.
The length in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for
CDS which show significant similarity to other CDS in the database.
The position of possible ribosome binding site sequences are given
where these have been used to deduce the initiation codon.
Gene prediction is based on positional base preference in codons
using a specially developed Hidden Markov Model (Krogh et al.,
Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot
program of Bibb et al., Gene 30:157-66(1984) as implemented at
<http://www.nih.go.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the
correct initiation codon. Where possible we choose an initiation
codon (atg, gtg, ttg or (att)) which is preceded by an upstream
ribosome binding site sequence (optimally 5-13bp before the
initiation codon). If this cannot be identified we choose the most
upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the
sequenced clone. It may be shorter because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring submissions.
Cosmid D95A lies between and overlaps cosmids D86A and D12A on the
AseI-D genomic restriction fragment.

FEATURES Location/Qualifiers

source 1. .45313
 /organism="Streptomyces coelicolor A3(2)"
 /strain="A3(2)"
 /db_xref="taxon:100226"
 /clone="cosmid D95A"
 gene 1. .79
 /gene="SCD95A.01"
 misc_feature 1. .105
 /note="nominal overlap with Streptomyces coelicolor cosmid StD86A"
 CDS <1. .79
 /gene="SCD95A.01"
 /note="SCD95A.01, unknown (fragment), len: >25 aa"
 /codon_start=2
 /transl_table=11
 /product="hypothetical protein SCD95A.01"
 /protein_id="CAB93028.1"
 /db_xref="GI:8248767"
 /translation="ISPTQLLDLAPVLAELGVRVLGEE"
 gene complement(91. .591)
 /gene="SCD95A.02c"
 CDS complement(91. .591)
 /gene="SCD95A.02c"
 /note="SCD95A.02c, unknown, len: 166 aa"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein SCD95A.02c"
 /protein_id="CAB93029.1"
 /db_xref="GI:8248768"
 /translation="MLDIGYALSTRFPDPPTDYRRADVHALRHDLFCGDVYLADTKA
 DRELSTAWGWVPVLDFAWALCDIVEQVDRDPAGSRAARPQRAELDFTESTDRLLFERR
 FGWVDIEAEWLPADEPPLSFSHTELRREARDFLHDLADLTDLHEDLADNPVWSLQA
 RFPRIP"
 gene complement(775. .1374)
 /gene="SCD95A.03c"
 CDS complement(775. .1374)
 /gene="SCD95A.03c"
 /note="SCD95A.03c, possible transcriptional regulator,
 len: 199 aa; similar to TR:CAB53122 (EMBL:AL109962)
 Streptomyces coelicolor putative transcriptional
 regulatory protein SCJ1.04, 207 aa; fasta scores: opt: 154
 z-score: 202.7 E(): 0.00079; 36.4% identity in 77 aa
 overlap. Contains Pfam match to entry PF00440 tetR,
 Bacterial regulatory proteins, tetR family"
 /codon_start=1
 /transl_table=11
 /product="putative transcriptional regulator"
 /protein_id="CAB93030.1"
 /db_xref="GI:8248769"
 /translation="MPDRTPDQPLTSRGAATHRRILDVATREFAEHGIAGARVERIVA
 AARTNKAQLYAYFGSKDGLFDAIFFGSLDRIVNVVPIDADDLADWAVRLYDEYLCRPD
 LIRLATWARLERRPAGHLVDDADRRDDAKLRAVAEAQAAGRVRPGDPFDVLALVIAMS
 MAWSPVSNVYAATADEPDDVHERRRALLRDAVRRATAPD"
 misc_feature complement(1174. .1314)
 /gene="SCD95A.03c"
 /note="Pfam match to entry PF00440 tetR, Bacterial
 regulatory proteins, tetR family, score 34.80, E-value

6.3e-08"

RBS complement(1375. .1378)

RBS 1480. .1485

gene 1490. .2626
/gene="SCD95A.04"

CDS 1490. .2626
/gene="SCD95A.04"
/note="SCD95A.04, probable NADP-dependent alcohol dehydrogenase, len: 378 aa; similar to SW:ADH_MYCTU (EMBL:AL021287) Mycobacterium tuberculosis NADP-dependent alcohol dehydrogenase (EC 1.1.1.2) Adh, 346 aa; fasta scores: opt: 986 z-score: 1074.5 E(): 0; 45.0% identity in 349 aa overlap. Contains Pfam match to entry PF00107 adh_zinc, Zinc-binding dehydrogenases and match to Prosite entry PS00059 Zinc-containing alcohol dehydrogenases signature"
/codon_start=1
/transl_table=11
/product="putative NADP-dependent alcohol dehydrogenase"
/protein_id="CAB93031.1"
/db_xref="GI:8248770"
/translation="MRTTVGWQATGPTLRRAPLERDLRPDDLAVRVVDYCGVCHTDLH AVRAGAADGSGPRPLVPGHEFTGVVTETGTAVTRFRPGDPVAVGNIVDSCGTCAMCRI GQENFCHSFPTLTYYGGTDRRDGSTTLGGYSREYVLGERFAYALPAALDPAAAPLLCA GITVWEPLRALGAGPGRVAVAGLGGLGHLAVKLAVALGADTSVISRSPDKAEDARRL GARELVVSTDPERLAAARERFDIVVDTVSAPHDLGAYLRLVALDGTLSHLGHLGPVTV ETLDLLIGRKKLSSAGSGGRKGTAEMLAFCAEHGITADIELLPSARVNEALDRLDRGD VRHRFVLDSLSDHSGPLDGLSDPSDPLRGPDRSGPQDPAAG"

misc_feature 1496. .2515
/gene="SCD95A.04"
/note="Pfam match to entry PF00107 adh_zinc, Zinc-binding dehydrogenases, score 299.10, E-value 5.5e-86"

misc_feature 1595. .1612
/gene="SCD95A.04"
/note="PS00190 Cytochrome c family heme-binding site signature"

misc_feature 1673. .1717
/gene="SCD95A.04"
/note="PS00059 Zinc-containing alcohol dehydrogenases signature"

gene 2636. .3607
/gene="SCD95A.05"

CDS 2636. .3607
/gene="SCD95A.05"
/note="SCD95A.05, possible mutT-like protein, len: 323 aa; similar to TR:O69888 (EMBL:AL023797) Streptomyces coelicolor hypothetical 19.4 kD protein SC2E1.17 or MutT, 172 aa; fasta scores: opt: 266 z-score: 322.5 E(): 1.7e-10; 39.3% identity in 145 aa overlap. Contains Pfam match to entry PF00293 mutT, Bacterial mutT protein and match to Prosite entry PS00893 mutT domain signature. High content in alanine and leucine amino acid residues"
/codon_start=1
/transl_table=11
/product="putative mutT-like protein"
/protein_id="CAB93032.1"
/db_xref="GI:8248771"

```

                                /translation="MILTPLALTPDHDIPGPVLTELTALYASHRAFHALSGDFDPED
                                IRPEQVATALADELARPGAEVLLARDAGRLVGAVTLARHPDPSDPDWIGLLMVDAG
                                LTRQGHGSRLVSLVEDRFRAAGRTAVRLAVLDGNTDALSFWTALGYRVLDHRRDLGAE
                                RPCTVLRRELASDKPRTPRRAARVAVLDPEGAVFLLRYDNVEVG VHWAMPGGGLEADE
                                NPREGALREVREETGWTDLPGPLLCTWEHDFTHLSVGPVRQYEHIIYVAQGP RREPTG
                                PHLAAHAADGILTWRWWSRAELAEAPEPLWPPDLALLLETFGGREG"
misc_feature 3212. .3337
                                /gene="SCD95A.05"
                                /note="Pfam match to entry PF00293 mutT, Bacterial mutT
                                protein, score 33.50, E-value 4.1e-08"
misc_feature 3269. .3328
                                /gene="SCD95A.05"
                                /note="PS00893 mutT domain signature"
gene          complement(3632. .5053)
                                /gene="SCD95A.06c"
CDS           complement(3632. .5053)
                                /gene="SCD95A.06c"

```

```

Query Match          10.1%; Score 41.4; DB 1; Length 45313;
Best Local Similarity 47.8%; Pred. No. 1.3;
Matches 120; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

```

```

Qy      90  acccatcccaacactccaaaccaacacacttcaaccaaaccaccacaacaatgccttc 149
          || |||||      | |  || |||  | || ||  | || |  | || |
Db  40284  ACTCATCCGGGACGCCTACCGCAGCCAGGCCGACAGCCTCCGCGCCGCCGCGTCCGG 40343

Qy     150  agtaaccaggcccggtctcatgtggcgtagcgtagcgccgcggcatggttagagccccaccc 209
          |      || |||| ||||| |      |||| | | | | | || | | | |
Db  40344  CGCCGACCTGGCCGGTCTCGCGCACGCCCTGCGCGCCTGGGCCCTCGACGACCCGCAGCG 40403

Qy     210  cttcgctcgcaatcccatcaccatgaccctcacgcctggcgcgccgacgtctccaa 269
          || |  | || | | | | | | | | | | | | | | | | | | | |
Db  40404  CTACTTCCTCATCTTCGGTACCCCGTCCCGGCTACCGGGCGCCCGACGACATCACCGA 40463

Qy     270  gaaagtcgtgaagacaagcactgtcttcttcccttctatgcaggtatccttggatggcc 329
          || | | | | |||| | | ||| || | | | | | | | | | | |
Db  40464  GATCGCCGCCGAGACCATGGCGGTCATCGTCGACGCCTGTGCCGCACTGCCTCCGTCGGA 40523

Qy     330  agtcgcagccg 340
          | | | | |
Db  40524  CGGCACCGACG 40534

```

```

RESULT      5
AC023519
LOCUS       AC023519      60609 bp      DNA           HTG           13-JUL-2000
DEFINITION  Homo sapiens clone RP11-16I22, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC023519
VERSION     AC023519.2   GI:7144934
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 60609)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens, clone RP11-16I22

```

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 60609)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavkiy,L.,
Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Largocque,K., Lehoczký,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M.,
Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
Zody,M.

TITLE Direct Submission
JOURNAL Submitted (15-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Mar 3, 2000 this sequence version replaced gi:6978163.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3549
Center clone name: 16_I_22

* NOTE: This record contains 69 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 779: contig of 779 bp in length
* 780 879: gap of 100 bp
* 880 1683: contig of 804 bp in length
* 1684 1783: gap of 100 bp
* 1784 2597: contig of 814 bp in length
* 2598 2697: gap of 100 bp
* 2698 3494: contig of 797 bp in length

* 3495 3594: gap of 100 bp
 * 3595 4365: contig of 771 bp in length
 * 4366 4465: gap of 100 bp
 * 4466 5248: contig of 783 bp in length
 * 5249 5348: gap of 100 bp
 * 5349 6130: contig of 782 bp in length
 * 6131 6230: gap of 100 bp
 * 6231 7007: contig of 777 bp in length
 * 7008 7107: gap of 100 bp
 * 7108 7872: contig of 765 bp in length
 * 7873 7972: gap of 100 bp
 * 7973 8748: contig of 776 bp in length
 * 8749 8848: gap of 100 bp
 * 8849 9624: contig of 776 bp in length
 * 9625 9724: gap of 100 bp
 * 9725 10485: contig of 761 bp in length
 * 10486 10585: gap of 100 bp
 * 10586 11356: contig of 771 bp in length
 * 11357 11456: gap of 100 bp
 * 11457 12270: contig of 814 bp in length
 * 12271 12370: gap of 100 bp
 * 12371 13158: contig of 788 bp in length
 * 13159 13258: gap of 100 bp
 * 13259 14034: contig of 776 bp in length
 * 14035 14134: gap of 100 bp
 * 14135 14923: contig of 789 bp in length
 * 14924 15023: gap of 100 bp
 * 15024 15789: contig of 766 bp in length
 * 15790 15889: gap of 100 bp
 * 15890 16659: contig of 770 bp in length
 * 16660 16759: gap of 100 bp
 * 16760 17531: contig of 772 bp in length
 * 17532 17631: gap of 100 bp
 * 17632 18415: contig of 784 bp in length
 * 18416 18515: gap of 100 bp
 * 18516 19322: contig of 807 bp in length
 * 19323 19422: gap of 100 bp
 * 19423 20177: contig of 755 bp in length
 * 20178 20277: gap of 100 bp
 * 20278 21084: contig of 807 bp in length
 * 21085 21184: gap of 100 bp
 * 21185 21938: contig of 754 bp in length
 * 21939 22038: gap of 100 bp
 * 22039 22759: contig of 721 bp in length
 * 22760 22859: gap of 100 bp
 * 22860 23617: contig of 758 bp in length
 * 23618 23717: gap of 100 bp
 * 23718 24487: contig of 770 bp in length
 * 24488 24587: gap of 100 bp
 * 24588 25407: contig of 820 bp in length
 * 25408 25507: gap of 100 bp
 * 25508 26337: contig of 830 bp in length
 * 26338 26437: gap of 100 bp
 * 26438 27252: contig of 815 bp in length
 * 27253 27352: gap of 100 bp
 * 27353 28104: contig of 752 bp in length
 * 28105 28204: gap of 100 bp

* 28205 28987: contig of 783 bp in length
 * 28988 29087: gap of 100 bp
 * 29088 29793: contig of 706 bp in length
 * 29794 29893: gap of 100 bp
 * 29894 30682: contig of 789 bp in length
 * 30683 30782: gap of 100 bp
 * 30783 31589: contig of 807 bp in length
 * 31590 31689: gap of 100 bp
 * 31690 32450: contig of 761 bp in length
 * 32451 32550: gap of 100 bp
 * 32551 33304: contig of 754 bp in length
 * 33305 33404: gap of 100 bp
 * 33405 34214: contig of 810 bp in length
 * 34215 34314: gap of 100 bp
 * 34315 35074: contig of 760 bp in length
 * 35075 35174: gap of 100 bp
 * 35175 35907: contig of 733 bp in length
 * 35908 36007: gap of 100 bp
 * 36008 36794: contig of 787 bp in length
 * 36795 36894: gap of 100 bp
 * 36895 37673: contig of 779 bp in length
 * 37674 37773: gap of 100 bp
 * 37774 38563: contig of 790 bp in length
 * 38564 38663: gap of 100 bp
 * 38664 39472: contig of 809 bp in length
 * 39473 39572: gap of 100 bp
 * 39573 40366: contig of 794 bp in length
 * 40367 40466: gap of 100 bp
 * 40467 41250: contig of 784 bp in length
 * 41251 41350: gap of 100 bp
 * 41351 42103: contig of 753 bp in length
 * 42104 42203: gap of 100 bp
 * 42204 42970: contig of 767 bp in length
 * 42971 43070: gap of 100 bp
 * 43071 43841: contig of 771 bp in length
 * 43842 43941: gap of 100 bp
 * 43942 44742: contig of 801 bp in length
 * 44743 44842: gap of 100 bp
 * 44843 45620: contig of 778 bp in length
 * 45621 45720: gap of 100 bp
 * 45721 46513: contig of 793 bp in length
 * 46514 46613: gap of 100 bp
 * 46614 47373: contig of 760 bp in length
 * 47374 47473: gap of 100 bp
 * 47474 48250: contig of 777 bp in length
 * 48251 48350: gap of 100 bp
 * 48351 49143: contig of 793 bp in length
 * 49144 49243: gap of 100 bp
 * 49244 49999: contig of 756 bp in length
 * 50000 50099: gap of 100 bp
 * 50100 50886: contig of 787 bp in length
 * 50887 50986: gap of 100 bp
 * 50987 51802: contig of 816 bp in length
 * 51803 51902: gap of 100 bp
 * 51903 52707: contig of 805 bp in length
 * 52708 52807: gap of 100 bp
 * 52808 53625: contig of 818 bp in length

* 53626 53725: gap of 100 bp
 * 53726 54485: contig of 760 bp in length
 * 54486 54585: gap of 100 bp
 * 54586 55372: contig of 787 bp in length
 * 55373 55472: gap of 100 bp
 * 55473 56229: contig of 757 bp in length
 * 56230 56329: gap of 100 bp
 * 56330 57101: contig of 772 bp in length
 * 57102 57201: gap of 100 bp
 * 57202 57982: contig of 781 bp in length
 * 57983 58082: gap of 100 bp
 * 58083 58857: contig of 775 bp in length
 * 58858 58957: gap of 100 bp
 * 58958 59739: contig of 782 bp in length
 * 59740 59839: gap of 100 bp
 * 59840 60609: contig of 770 bp in length.

FEATURES Location/Qualifiers

Query Match 10.0%; Score 41.2; DB 2; Length 60609;
 Best Local Similarity 50.3%; Pred. No. 1.5;
 Matches 94; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

Qy 82 accaatacacccatcccaacactccaaaccaacacacttcaaccaaaccaccacaaca 141
 ||| | ||| || ||| || | | || ||| | | || ||| | | | | ||
 Db 49656 ACCCACACAACCCACCCACCCACCCACCCACCCACCCACCCACCCACCCACCCACCCCA 49715
 Qy 142 atgccttcagtaaccaggcccggtctcatgtggcgtagcgtagcgccgcggcatggtagag 201
 | || | || || ||| || | | | | | || | || | || |
 Db 49716 ACCCCCCCTCCCACACACNCCCCCCCCCCCCACCACCCCNCAACCCCNACACCCACCA 49775
 Qy 202 cccacccccttcgctcgcaatcccatcaccatgacccctcacgcctggcgcgccgcccac 261
 ||||| || | ||| ||| ||| ||| || | | || || || |
 Db 49776 CCCACCCCCCACCTANCNCCCCACCACCCTACCCCCAAACCACCCACACCCCCCCC 49835
 Qy 262 ctctcca 268
 | | |||
 Db 49836 CCCACCA 49842

RESULT 6
 AP004043/c
 LOCUS AP004043 102242 bp DNA HTG 17-AUG-2001
 DEFINITION Oryza sativa chromosome 2 clone OJ1124_D06, *** SEQUENCING IN
 PROGRESS ***, in ordered pieces.
 ACCESSION AP004043
 VERSION AP004043.1 GI:15208411
 KEYWORDS HTG; HTGS_PHASE2.
 SOURCE Oryza sativa (cultivar:Nipponbare) DNA, clone:OJ1124_D06.
 ORGANISM Oryza sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzeae; Oryza.
 REFERENCE 1 (bases 1 to 102242)
 AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.
 TITLE Oryza sativa nipponbare(GA3) genomic DNA, chromosome 2, BAC
 clone:OJ1124_D06
 JOURNAL Published Only in Database (2001) In press

REFERENCE 2 (bases 1 to 102242)
AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE Direct Submission
JOURNAL Submitted (15-AUG-2001) Takuji Sasaki, National Institute of
Agrobiological Resources, Rice Genome Research Program; Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT The nucleotide sequence of this BAC clone was generated by
combining Monsanto and RGP-Japan sequencing data.
NOTE: It currently consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces is believed
to be correct as given, however the sizes of the gaps between them
are based on estimates that have provided by the submitter. This
sequence will be replaced by the finished sequence as soon as it is
available and the accession number will be preserved.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES Location/Qualifiers
source 1. .102242
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="2"
/clone="OJ1124_D06"
BASE COUNT 29287 a 22590 c 22673 g 27691 t 1 others
ORIGIN

Query Match 9.9%; Score 40.8; DB 2; Length 102242;
Best Local Similarity 49.5%; Pred. No. 1.8;
Matches 105; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

Qy 89 caccatcccaacactccaaaccaaccaacttcaaccaaaccaccacaacaatgcctt 148
| | | | | | | | | | | | | | | | | | | | | |
Db 55012 CTCTCCTCTCCGCCTACACGACAACCGCCTCTACGACCGCGCCATCCAAGCCTTCCGCA 54953
Qy 149 cagtaaccagggcccggtctcatgtggcgtagcgctcgcccgcgccatggttagagccccacc 208
| | | | | | | | | | | | | | | | | | | | | |
Db 54952 CTCTCCCGCGCGAGCTCGGCATCAAGCCCAGCGTCGTCTCTCACAACGTCCTTCTCAAGT 54893
Qy 209 ccttcgctcgcaatcccatcaccatgacccctcacgcctggcgcgccgcccacctctcca 268
| | | | | | | | | | | | | | | | | | | | | |
Db 54892 CCTTTGTTGCCAGTGGCGACCTCGCCTCCGCCCGCGCCCTGTTTCGATGAAATGCCTTCCA 54833
Qy 269 agaaagtcgtgaagacaagcactgtcttcttc 300
| | | | | | | | | | | | | | | | | | | | | |
Db 54832 AGGCTGACGTCGAGCCAGACATTGTCTCCTGC 54801

RESULT 7
AP003865/c
LOCUS AP003865 125118 bp DNA HTG 10-JUL-2001
DEFINITION Oryza sativa chromosome 8 clone OJ1081_B12, *** SEQUENCING IN
PROGRESS ***, in ordered pieces.

Db 37735 CCCCNCCCCCCCCCNCCCCCNCCCCCCCCCCCCCNCCNCCCCCCCCCNCCNCCCCCCCC 37676

```

      | | |
Db  37675 CCCCCC 37670

```

AF233344/c

ACCESSION AF233344

KEYWORDS

ORGANISM

REFERENCE 1 (bases 1 to 3338)

TITLE Tumour suppressive properties of fibroblast growth factor receptor 2-IIIb in human bladder cancer

MEDLINE 20071102

AUTHORS Girault, J., Radvanyi, F. and Ricol, D.

JOURNAL Submitted (09-FEB-2000) UMR 144, CNRS-Institut Curie, 26 rue d'Ulm,
Paris 75248, France

```
source      1. .3338  
            /organism="Homo sapiens"  
            /db_xref="taxon:9606"  
            /chromosome="10"  
            /map="10q26"  
            /clone="PAC6539"
```

```

/ gene="FGFR2"

```

```

/gene="FGFR2"
/note="fibroblast growth factor receptor 2"

```

ORIGIN

Best Local Similarity 55.3%; Pred. No. 2.9;

Matches 78; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

Qy 127 caaaccaccacaacaatgccttcagtaacccaggcccgtctcatgtggcgtagcgtcgcc 186

Db 3207 CCAGCCCGGAGAGCAGTCGCCGCGCCGGGCCAGGTACGCCGCATGCAGCCCCGCGGCGCC 3148

Qy 187 cgcgcatggttagagccccacccttcgctcgcaatcccatcaccatgaccctcagcc 246
 || | || ||| | | | || ||| | | ||| ||| | |||
 Db 3147 CGAGCTTTGTGGCGGCCGCGCGCTCCCTCGCCGCTCCGCACCCGCCGCCGCCGCT 3088
 Qy 247 tggcgcgccgcccacctctcc 267
 ||| | || ||| |||||
 Db 3087 CGGCTCTCCACCGCGCTCTCC 3067

RESULT 9
 AC009988/c
 LOCUS AC009988 179004 bp DNA HTG 05-MAY-2001
 DEFINITION Homo sapiens chromosome 10 clone RP11-62L18, WORKING DRAFT
 SEQUENCE, 9 unordered pieces.
 ACCESSION AC009988
 VERSION AC009988.10 GI:13957605
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 179004)
 AUTHORS Smith,D.R.
 TITLE Genome Therapeutics Corporation Sequencing Center: Human Genome
 Sequence Data
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 179004)
 AUTHORS Smith,D.R.
 TITLE Direct Submission
 JOURNAL Submitted (09-SEP-1999) Genome Therapeutics Corporation, 100 Beaver
 Street, Waltham, MA 02453, USA
 COMMENT On May 5, 2001 this sequence version replaced gi:13605974.
 ----- Genome Center
 Center: Genome Therapeutics Corporation
 Center code: GTC
 Web site: <http://www.genomecorp.com/>
 Contact: gtc-seqcenter@genomecorp.com
 ----- Project Information
 Center project name: hg002
 ----- Summary Statistics
 Sequencing vector: N/A
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 990315
 Consensus quality: 173303 bases at least Q40
 Consensus quality: 174498 bases at least Q30
 Consensus quality: 175282 bases at least Q20
 Insert size: 178303; sum-of-contigs
 Quality coverage: 3.9x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 9 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

*      1      1196: contig of 1196 bp in length
*    1197      1296: gap of unknown length
*    1297      2376: contig of 1080 bp in length
*    2377      2476: gap of unknown length
*    2477      3501: contig of 1025 bp in length
*    3502      3601: gap of unknown length
*    3602      4746: contig of 1145 bp in length
*    4747      4846: gap of unknown length
*    4847      6274: contig of 1428 bp in length
*    6275      6374: gap of unknown length
*    6375     15038: contig of 8664 bp in length
*   15039     15138: gap of unknown length
*   15139     42629: contig of 27491 bp in length
*   42630     42729: gap of unknown length
*   42730     98264: contig of 55535 bp in length
*   98265     98364: gap of unknown length
*   98365    179004: contig of 80640 bp in length.

```

```

FEATURES             Location/Qualifiers
     source            1. .179004
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="10"
                        /clone="RP11-62L18"
                        /clone_lib="RPCI-11"
     misc_feature      1. .1196
                        /note="assembly_name:Contig1"
     misc_feature      1297. .2376
                        /note="assembly_name:Contig2"
     misc_feature      2477. .3501
                        /note="assembly_name:Contig3"
     misc_feature      3602. .4746
                        /note="assembly_name:Contig4"
     misc_feature      4847. .6274
                        /note="assembly_name:Contig7
                        clone_end:T7"
     misc_feature      6375. .15038
                        /note="assembly_name:Contig8
                        clone_end:SP6"
     misc_feature      15139. .42629
                        /note="assembly_name:Contig9"
     misc_feature      42730. .98264
                        /note="assembly_name:Contig10"
     misc_feature      98365. .179004
                        /note="assembly_name:Contig11"
BASE COUNT    47825 a  41271 c  40632 g  48463 t    813 others
ORIGIN

```

```

Query Match          9.8%;  Score 40.2;  DB 2;  Length 179004;
Best Local Similarity 55.3%;  Pred. No. 2.6;
Matches 78;  Conservative 0;  Mismatches 63;  Indels 0;  Gaps 0;

```

```

Qy   127 caaaccaccacaacaatgccttcagtaaccaggcccggtctcatgtggcgtagcgctgcc 186
      ||||  |||||  |  |  |||||  ||  ||||  ||  |||||
Db  46786 CCAGCCCGGAGAGCAGTCGCCGCGCCGGGCCAGGTACGCCGCATGCAGCCCCGCGGCGCC 46727

Qy   187 cgcggcatggtagagccccacccttcgctcgcaatcccatcaccatgaccctcagcc 246

```


Db 46726 CGAGCTTTGTGGCGGCCGCGCGCTCCCTCGCCCGCTCCGCACCCGCCGCCGCCGCT 46667
 Qy 247 tggcgcgccgcccacctctcc 267
 Db 46666 CGGCTCTCCACCGCGCTCTCC 46646

RESULT 10

AF275943

LOCUS AF275943 11096 bp DNA BCT 02-SEP-2000

DEFINITION Streptomyces avermitilis avermectin polyketide synthase gene, partial cds.

ACCESSION AF275943

VERSION AF275943.1 GI:9964075

KEYWORDS .

SOURCE Streptomyces avermitilis.

ORGANISM Streptomyces avermitilis

Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE 1 (bases 1 to 11096)

AUTHORS Hong, Y.-S. and Lee, J.J.

TITLE Targeted Gene Disruption of the avermectin O-methyltransferase gene and polyketide synthase gene from Streptomyces avermitilis

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 11096)

AUTHORS Hong, Y.-S. and Lee, J.J.

TITLE Direct Submission

JOURNAL Submitted (07-JUN-2000) Anticancer Agent Research Laboratory, Korea Research Institute of Bioscience and Biotechnology, P.O. Box 116, YuSong-Gu, Taejon 305-600, South Korea

FEATURES Location/Qualifiers

source

1. .11096

/organism="Streptomyces avermitilis"

/strain="ATCC31271"

/db_xref="taxon:33903"

/db_xref="ATCC:31271"

CDS

220. .>11096

/note="Avr1"

/codon_start=1

/transl_table=11

/product="avermectin polyketide synthase"

/protein_id="AAG09812.1"

/db_xref="GI:9964076"

/translation="MADEADGGVVFVFPQGQWPQWPGMGRELLDASDVFRSVRACEAE
FAPYVDWSVEQVLRDSPDAPGLDRVDVVQRTLFVAVMISLAALWRSQGVEPCAVLGHS
GEIAAAHVSGGLSLADAARVGDAWSQAQTTLAGTGALVSAATPELLPRIAPWTE
PARLAVAAVNGPRSTVVSGAREAVADLVADLTAAQVRTRMIPVDVPAHSPLMYAIEER
VVSGLLPITPRPSRIPFHSSVTGGRLDTRDLAAYWYRNMSSTVRFEPARLLQLQGP
KTFVEMSPHPVLTMLGLQELADLGDTTGTADTVIMGTLRRGGTLDHFLTSLAQLRGH
GETSATTVLSARLTALSPTQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHL
GFDSVMGVLELRNRLSKATGLRLPVTLILDHTTAAVAARLRTAALGHLEDTAPVPDS
PSGHGGTAAADDPIAIIIGMACRFPGGVRSPKDLWELAASGGDAIGPFPTDRGWPTQR
HAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGISPREALAMPQORLLLETWEAFERA
GIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEGAADVEGHILTGTGTVLSGRSAYS
FGLEGPAITVDTGCSASLVTLHLACQSLRSGETFALAGGVSDVHPGMFIEFSRQCGL
SVDGRCKAVSAAADGTGWGEGIGLLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQAL

ANAGLSVADVDDVEGHGTGTTLGDPIDAEQALLATYQQRAGDRPLWLGLSKSNIGHTMA
AAGVGGVIKMMALREGVLPRTLHVDEPSQGLAAGAVRLLTEAVPWPFGDAAGRLRRA
GVSSFGIGGTNAHVILEEAPAAGGCVAGGRVLEGAPGLAISVAESVAAPVAVSAPVAE
SVPVPVPVPVPVPVSARSEAGLRAQAEALRQYEAVQPDVSLADVGAGLACRQAVLEHH
VVILAACTSSSRSAARTTARSSSTARPQARPAPTSARRIGALAAGSGSAALTGTHAPG
GDRGGVVFVFPWQGGQWAGMGVRLCLLRVFAARMQACEEALAPWVDWSVVDILRDA
GDAVWEQADVVQPVLFVSMVSLAALWRSYGIENEVLGHSKDEIAAAHIYGALSLKDA
AKTVALPPQEVEQLIGERGGRLWVAAVNGPRSTAVSGDAEAVVEVLAYCAGTGVRPRI
PVDYASHCPHVQPLREELLELLGDISPQPYGVFFSTVEGTWLDTTTTLDAAYWYRNLH
QPVRFSHDVQALADQGHRVLEVSHPHTLVPAIEDTTEDTARRRHCDRQPPRRERHP
LLPQRLRLDPYYRHRQTHHVAPLLDPPRHLPPPLDAPRPAHLPFQHQHYWLESSQPGA
SGSGGAGAGSGAGSGRAGTAGGTAEVESRFWDVARQDLETVATTAVPPSAGLDTVV
PALSAWHRHQHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAPETQTHHPHITNIL
TNLHHHGITPIPLPLTTHHTNPQHLHHTLHHTRQQAQNHTTGAITGLLSLLALDETPH
PHHPHTPTGTLLNLTLTQHTQTHPPTPLWYATTNATTTHPNDPLTHQTQAQTWGLAR
TTLLEHPTHTAGIIDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGHTRRRLPTTL
TPTHQPPTPTPHGTTTLITGGTGALATHLTHHLTTHQPTQHLLLTSTRTGPHTPHAQHLT
TQLQQKGIHLTITTCDDTSTPRPTHNNSLNTIPPQHPVTTVIHTGGILDDATLTNLPTPT
QLNNVLRKAHSAHLLHQLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAH
HRHTHHLPATSIWGTWQGNGLADSDKARAYLDRRGFRMSPPELATAAVTQAIADTER
PYVVIADIDWSKIEHTSQTSDLVSAAREREPAPVQRPTPPAELHKTLAHQTSADQRAAL
LELVDRDHVAAVLRHADPKAIAPDQSFALGFDSLTAVEFRNLLIKATGLRLPVSLVFD
HPTPAKLAVHLQNLRGTAESAAPSAAAVTAEASVTEPIAIVGMACRFPGGVTSADDF
WDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGIS
PREALAMPQQRLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDYALRVHNAGQS
TDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVALHLACQALRAGECSMAL
AGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAADGTGWGEGVGMLLVERLSDAHRN
GHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAGDVADEAHGTGT
TLGDPIDAEQALLATYQQRAGEGRLWLGLSVKSNVGHGTQAAAGVAGVIKMMALRHGLL
PRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEA
PADDDVPGGPPAGEGDAGSDDEAAAGSPGVWPWLVSQALRAQAQALHAHLTDHPG
LDLADVGYTTLAARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVHSSAPGGTGT
GEAAGKTAFICSGQGTQRPMAHGLYHTHPVFAALNDICTHLDPHLDPLPLLTQN
DNDNEDAAALLQQTRYAQPALFAFQVALHRLTLDGYHITPHYHAGHSLGEITGAHLAG
ILTLTDATTLITQRATLMQTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSL
VISGTPHTVQHITTLCCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPHT
PLITANTPPDQLLTPHYWTQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHH
NLPNTPTTTLTLTHPHHPQTHLLTNLAKTTTTWHPHHYTHHAQTPHPHPLDLPTYP
FQHHHYWLESTQPGAGNVSAAGLDPTHEPPLGATLELATDGGALLAGRLSLRSHPWLA
DHAVGGTVLLSGATFLELALHAGTYVGC DRVDELTLHAPLVVPVDDGGVSVQVGAAAD
GKGRRLVSVYARGGSACGGGGASGGVWTCHASGVLVEAAAGGVVDGLAGVWPPRGAV
AVDVGVRDRLAGAGCVLGPVFSGLRAVWRDGGDLAEVCLPEEAWGDAAGFGLHPAL
LNGVVQPLSVLLPGGTGFGKAGFGKGVVPAVWGGVSLHRAGVTGVRVRVSAVGRGG
GREAVSVVVGDEAGVPVASVDRLELRPVDMDQLRAVSVSAGRRGSLYAVQWAEVGPVP
VCGQAWAWHEDVGESGGGPVPGVVVLRCPDAGAGGGGGGGGGGGVGEVGGVGLGVVQG
WLGLERFAGSRLVVVTRGAVVAGPEDGPDVVGASVWGLVRSQAQAEHPDRFVLLDLDT
DTGTDLDTGAGAGWGDGGRVAAVVACGEPLAVRGERLLAARLTRLESSGDVPAQRS
GDTRARRSDVPAQRSGGVPAARRSVDVSGREVLPLWLSGGSVLVTGGTGVLGAAVARHLA
GVCGRDLLLLVSRGPDAPGAEGLRRSWPRGAEVRIACDVGERREVVRLAGGCSCRV
SVFVDFSAVASVTVRLPVASDVRKEAAMAYATVEEFTDYLDPPD"

BASE COUNT 1756 a 4059 c 3528 g 1753 t
ORIGIN

Query Match 9.7%; Score 39.8; DB 1; Length 11096;
Best Local Similarity 46.8%; Pred. No. 3.6;
Matches 125; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

Qy 77 gcgataccaatacacccatcccaacactccaaaccaaccaacacttcaaccaaaccacca 136
 |||| |||| ||||| |||| | | | | | | | | | |
 Db 5441 GCGACACCAGCACACCCAGACCAACTCACAACAACTCACTCAACACCATCCCCCACAAC 5500

Qy 137 caacaatgccttcagtaacccaggcccggtctcatgtggcgtagcgctcgcccgcgccatgg 196
 | | | | | | | | | | | | | | | | | | | |
 Db 5501 ACCCCGTCACCACCGTCATCCACACCGGAGGCATCCTCGACGACGCCACCCTCACCAACC 5560

Qy 197 tagagccccaccccttcgctcgcaatcccatcaccatgacccctcacgcctggcgcgccg 256
 | ||| || || | | | | | | | | | | | | | |
 Db 5561 TCACCCCAACCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCACCTCCTCC 5620

Qy 257 ccgacctctccaagaaagtcgtgaagacaagcactgtcttcttcccttctatgcaggta 316
 | | ||| || | | | | | | | | | | | | | |
 Db 5621 ACCAACTCACCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCA 5680

Qy 317 tccttgatggccagtcgcagccgcct 343
 | | | | | | | | | |
 Db 5681 CCTTCGGCGCACCCGGCCAAGCCAACT 5707

RESULT 11

AC083782/c

LOCUS AC083782 56494 bp DNA HTG 30-SEP-2000

DEFINITION Homo sapiens chromosome 12 clone RP11-84L9 map 12, LOW-PASS
 SEQUENCE SAMPLING.

ACCESSION AC083782

VERSION AC083782.1 GI:10440689

KEYWORDS HTG; HTGS_PHASE0.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 56494)

AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE Homo sapiens chromosome 12, clone RP11-84L9

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 56494)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L.,
 Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
 Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
 DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
 FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
 Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
 Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,
 Lamazares,R., Landers,T., Lehoczký,J., Levine,R., Lieu,C., Liu,G.,
 Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
 McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
 Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
 O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K.,
 Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
 Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
 Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,

Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (30-SEP-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L11227

Center clone name: 84_L_9

* NOTE: This record contains 72 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

* 1 634: contig of 634 bp in length
* 635 734: gap of 100 bp
* 735 1430: contig of 696 bp in length
* 1431 1530: gap of 100 bp
* 1531 2223: contig of 693 bp in length
* 2224 2323: gap of 100 bp
* 2324 3001: contig of 678 bp in length
* 3002 3101: gap of 100 bp
* 3102 3794: contig of 693 bp in length
* 3795 3894: gap of 100 bp
* 3895 4593: contig of 699 bp in length
* 4594 4693: gap of 100 bp
* 4694 5386: contig of 693 bp in length
* 5387 5486: gap of 100 bp
* 5487 6141: contig of 655 bp in length
* 6142 6241: gap of 100 bp
* 6242 6945: contig of 704 bp in length
* 6946 7045: gap of 100 bp
* 7046 7653: contig of 608 bp in length
* 7654 7753: gap of 100 bp
* 7754 8449: contig of 696 bp in length
* 8450 8549: gap of 100 bp
* 8550 9247: contig of 698 bp in length
* 9248 9347: gap of 100 bp
* 9348 10038: contig of 691 bp in length
* 10039 10138: gap of 100 bp
* 10139 10836: contig of 698 bp in length
* 10837 10936: gap of 100 bp
* 10937 11647: contig of 711 bp in length

* 11648 11747: gap of 100 bp
 * 11748 12433: contig of 686 bp in length
 * 12434 12533: gap of 100 bp
 * 12534 13240: contig of 707 bp in length
 * 13241 13340: gap of 100 bp
 * 13341 14065: contig of 725 bp in length
 * 14066 14165: gap of 100 bp
 * 14166 14877: contig of 712 bp in length
 * 14878 14977: gap of 100 bp
 * 14978 15673: contig of 696 bp in length
 * 15674 15773: gap of 100 bp
 * 15774 16374: contig of 601 bp in length
 * 16375 16474: gap of 100 bp
 * 16475 17104: contig of 630 bp in length
 * 17105 17204: gap of 100 bp
 * 17205 17906: contig of 702 bp in length
 * 17907 18006: gap of 100 bp
 * 18007 18692: contig of 686 bp in length
 * 18693 18792: gap of 100 bp
 * 18793 19481: contig of 689 bp in length
 * 19482 19581: gap of 100 bp
 * 19582 20296: contig of 715 bp in length
 * 20297 20396: gap of 100 bp
 * 20397 21077: contig of 681 bp in length
 * 21078 21177: gap of 100 bp
 * 21178 21915: contig of 738 bp in length
 * 21916 22015: gap of 100 bp
 * 22016 22700: contig of 685 bp in length
 * 22701 22800: gap of 100 bp
 * 22801 23507: contig of 707 bp in length
 * 23508 23607: gap of 100 bp
 * 23608 24314: contig of 707 bp in length
 * 24315 24414: gap of 100 bp
 * 24415 25032: contig of 618 bp in length
 * 25033 25132: gap of 100 bp
 * 25133 25833: contig of 701 bp in length
 * 25834 25933: gap of 100 bp
 * 25934 26637: contig of 704 bp in length
 * 26638 26737: gap of 100 bp
 * 26738 27443: contig of 706 bp in length
 * 27444 27543: gap of 100 bp
 * 27544 28159: contig of 616 bp in length
 * 28160 28259: gap of 100 bp
 * 28260 28961: contig of 702 bp in length
 * 28962 29061: gap of 100 bp
 * 29062 29783: contig of 722 bp in length
 * 29784 29883: gap of 100 bp
 * 29884 30482: contig of 599 bp in length
 * 30483 30582: gap of 100 bp
 * 30583 31206: contig of 624 bp in length
 * 31207 31306: gap of 100 bp
 * 31307 31931: contig of 625 bp in length
 * 31932 32031: gap of 100 bp
 * 32032 32719: contig of 688 bp in length
 * 32720 32819: gap of 100 bp
 * 32820 33512: contig of 693 bp in length
 * 33513 33612: gap of 100 bp

* 33613 34332: contig of 720 bp in length
 * 34333 34432: gap of 100 bp
 * 34433 35123: contig of 691 bp in length
 * 35124 35223: gap of 100 bp
 * 35224 35968: contig of 745 bp in length
 * 35969 36068: gap of 100 bp
 * 36069 36747: contig of 679 bp in length
 * 36748 36847: gap of 100 bp
 * 36848 37521: contig of 674 bp in length
 * 37522 37621: gap of 100 bp
 * 37622 38331: contig of 710 bp in length
 * 38332 38431: gap of 100 bp
 * 38432 39140: contig of 709 bp in length
 * 39141 39240: gap of 100 bp
 * 39241 39926: contig of 686 bp in length
 * 39927 40026: gap of 100 bp
 * 40027 40718: contig of 692 bp in length
 * 40719 40818: gap of 100 bp
 * 40819 41503: contig of 685 bp in length
 * 41504 41603: gap of 100 bp
 * 41604 42310: contig of 707 bp in length
 * 42311 42410: gap of 100 bp
 * 42411 43121: contig of 711 bp in length
 * 43122 43221: gap of 100 bp
 * 43222 43938: contig of 717 bp in length
 * 43939 44038: gap of 100 bp
 * 44039 44715: contig of 677 bp in length
 * 44716 44815: gap of 100 bp
 * 44816 45525: contig of 710 bp in length
 * 45526 45625: gap of 100 bp
 * 45626 46244: contig of 619 bp in length
 * 46245 46344: gap of 100 bp
 * 46345 47044: contig of 700 bp in length
 * 47045 47144: gap of 100 bp
 * 47145 47842: contig of 698 bp in length
 * 47843 47942: gap of 100 bp
 * 47943 48554: contig of 612 bp in length
 * 48555 48654: gap of 100 bp
 * 48655 49340: contig of 686 bp in length
 * 49341 49440: gap of 100 bp
 * 49441 50147: contig of 707 bp in length
 * 50148 50247: gap of 100 bp
 * 50248 50970: contig of 723 bp in length
 * 50971 51070: gap of 100 bp
 * 51071 51781: contig of 711 bp in length
 * 51782 51881: gap of 100 bp
 * 51882 52572: contig of 691 bp in length
 * 52573 52672: gap of 100 bp
 * 52673 53370: contig of 698 bp in length
 * 53371 53470: gap of 100 bp
 * 53471 54081: contig of 611 bp in length
 * 54082 54181: gap of 100 bp
 * 54182 54883: contig of 702 bp in length

Query Match 9.7%; Score 39.8; DB 2; Length 56494;
 Best Local Similarity 58.6%; Pred. No. 3.4;
 Matches 68; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

Qy 61 atttgagttcatcctagcgataccaatacacccatcccaaacactccaaaccaaccaacac 120
 |||| |||| | | || ||| || |||| ||| | | | ||| ||| |
 Db 31161 ATTTCCAACCATCCCAACCATTCATCCATCCATTCATTCATTCATTCATTCATCCA 31102

Qy 121 ttcaaccaaaccaccacaacaatgccttcagtaacccaggcccggtctcatgtggcg 176
 | || ||| ||| || | |||| | || | ||| | | |||| |||
 Db 31101 TGCATCCATTCCATTCCATCCATGCATGCATCCATCCATACATNTATCATTGAGCG 31046

RESULT 12

SC6G10/c

LOCUS SC6G10 36734 bp DNA BCT 24-MAR-1999

DEFINITION Streptomyces coelicolor cosmid 6G10.

ACCESSION AL049497

VERSION AL049497.1 GI:4539196

KEYWORDS aminotransferase; cox1; cox2; cox3; cytochrome b; cytochrome c
 oxidase; gene duplication; glycosyl transferase; heme-binding;
 integral membrane protein; long chain fatty acid coA ligase;
 membrane transporter; phosphoribosylanthranilate transferase; qcrA;
 qcrB; qcrC; quinolinate synthetase; Rieske iron-sulfur protein;
 secreted protein; transcriptional regulator; trpD; two component
 sensor kinase/response regulator.

SOURCE Streptomyces coelicolor A3(2).

ORGANISM Streptomyces coelicolor A3(2)
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE 1 (bases 1 to 36734)

AUTHORS Redenbach, M., Kieser, H.M., Denapait, D., Eichner, A., Cullum, J.,
 Kinashi, H. and Hopwood, D.A.

TITLE A set of ordered cosmids and a detailed genetic and physical map
 for the 8 Mb Streptomyces coelicolor A3(2) chromosome

JOURNAL Mol. Microbiol. 21 (1), 77-96 (1996)

MEDLINE 97000351

REFERENCE 2 (bases 1 to 36734)

AUTHORS Seeger, K. and Harris, D.

JOURNAL Unpublished

REFERENCE 3 (bases 1 to 36734)

AUTHORS Bentley, S.D., Parkhill, J., Barrell, B.G. and Rajandream, M.A.

TITLE Direct Submission

JOURNAL Submitted (23-MAR-1999) Streptomyces coelicolor sequencing project,
 Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
 CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
 David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
 Colney, Norwich, Norfolk NR4 7UH, UK

COMMENT

Notes:

Streptomyces coelicolor sequencing at The Sanger Centre is funded
 by the BBSRC and Beowulf Genomics

Details of S. coelicolor sequencing at the Sanger Centre are
 available on the World Wide Web.

(URL; http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are
 numbered using the following system eg SC7B7.01c. SC (S.
 coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
 strand).

The more significant matches with motifs in the PROSITE database
 are also included but some of these may be fortuitous. The length
 in codons is given for each CDS.

Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.go.jp/>

[jun/cgi-bin/frameplot.pl](http://www.nih.go.jp/jun/cgi-bin/frameplot.pl). CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid 6G10 Lies between and overlaps with cosmids 6E10 and 5F7 on the AseI-C genomic restriction fragment. .

FEATURES	Location/Qualifiers
source	1..36734 /organism="Streptomyces coelicolor A3(2)" /strain="A3(2)" /db_xref="taxon:100226" /clone="cosmid 6G10"
gene	complement(1..120) /gene="SC6G10.01c"
CDS	complement(<1..120) /gene="SC6G10.01c" /note="SC6G10.01c, partial CDS, unknown, len: >40aa" /codon_start=1 /transl_table=11 /label=SC6G10.01c /product="hypothetical protein" /protein_id="CAB39855.1" /db_xref="GI:4539197" /translation="MARAGRSGTPHRRANPPPHRHPTADAPVPFTPMRTLLI"
misc_feature	1..103 /note="Nominal overlap with cosmid 6E10."
gene	complement(232..666) /gene="SC6G10.02c"
CDS	complement(232..666) /gene="SC6G10.02c" /note="SC6G10.02c, unknown, len: 144aa; similar to hypothetical proteins from Mycobacterium eg. TR:053519 (EMBL:AL021957) hypothetical protein from Mycobacterium tuberculosis (144 aa) fasta scores; opt: 479, z-score: 620.2, E(): 3.1e-27, (49.0% identity in 143 aa overlap)." /codon_start=1 /transl_table=11 /label=SC6G10.02c /product="hypothetical protein" /protein_id="CAB39856.1" /db_xref="GI:4539198" /translation="MAEHTSSSITIEAAPADVMAVIADFARYPDWTGEVKEAQVLATD"

EQGRAEQVRLVMDAGAIKDDQTLGYTWTGEHEVSWTLVKSQMLRSLDGSYLLRPAGTG
 TEVTYRLTVDVKIPMLGMIKRKAEKVIIDRALAGLKKRVESK"

gene complement(804. .1601)
 /gene="SC6G10.03c"

CDS complement(804. .1601)
 /gene="SC6G10.03c"
 /note="SC6G10.03c, unknown, len: 265aa; weak similarity to
 many eg. TR:P95860 (EMBL:Y08256) hypothetical protein (309
 aa) fasta scores; opt: 152, z-score: 186.9, E(): 0.0041,
 (24.8% identity in 270 aa overlap)."
 /codon_start=1
 /transl_table=11
 /label=SC6G10.03c
 /product="hypothetical protein"
 /protein_id="CAB39857.1"
 /db_xref="GI:4539199"
 /translation="MAPTPPRNTSTRVHVSDVHGNARDLARAGDGADALICLDLVL
 FLDYADHSRGIFPDLFGVANADRIVALRTARRFEEAREFGRRWLAEAGGEPRELIERA
 VRKQYAELEFAAFPTPTYATYGNVDVPPPLWPEYAGPGTTVLDGERVEIGGRVFGFVGGG
 LRTPMNTPEYISDEEYAAKIEAVGEVDVLCCHIPPEVPELVYDTPARRFERGSRALLD
 AIRRTRPRYALFGHVHQPLVRRMRVGATECVNVGHFASSGRPWALEW"

gene 1912. .3708
 /gene="SC6G10.04"

CDS 1912. .3708
 /gene="SC6G10.04"
 /note="SC6G10.04, probable long chain fatty acid coA
 ligase, len: 598aa; similar to many eg. TR:E1359128
 (EMBL:AL034443) putative long chain fatty acid coA ligase
 from Streptomyces coelicolor (608 aa) fasta scores; opt:
 1347, z-score: 1506.7, E(): 0, (51.3% identity in 608 aa
 overlap) and SW:LCFB_RAT long chain fatty acid coA ligase
 from Rattus norvegicus (Rat) (699 aa) fasta scores; opt:
 568, z-score: 634.7, E(): 4.7e-28, (30.2% identity in 589
 aa overlap). Contains Pfam match to entry PF00501
 AMP-binding, AMP-binding enzyme and Prosite match to
 PS00455 Putative AMP-binding domain signature."
 /codon_start=1
 /transl_table=11
 /label=SC6G10.04
 /product="putative long chain fatty acid coA ligase"
 /protein_id="CAB39858.1"
 /db_xref="GI:4539200"
 /translation="MREFSLPALYEV PADGNLT DIVRRNAAQHPDVAVIARKVGGVWQ
 DVTARAFLAEVHSAAKGLIASGVQPGDRVGLMSRTRYEWTL LDFAIWSAGAITVPVYE
 TSSPEQVQWILGDSGATACVVESAGHAAAVESVREQLPALKNVWQIDAGAVEELGRLG
 QDVTDR TVEERGSIAKADDPATIVYTS GTTGRPKGCVLTHRSFFAE CGNVVERLRPLF
 RTGECSVLLFLPLAHVFGRLVQVAPMIAPIKLG NVPDIKNLTDELA AFRPTLILGVPR
 VFEKVYNSARAKAQADGKGKIFDKAADTAIAYS KALDAPSGSPVGLKIKHKVFDKLVY
 SKLRTVLGGRGEY AISGGAPLGERLGHFFRGIGFTVLEGYGLTESCAATAFNPWDRQK
 IGTVGQPLPGSVVRIADDGEVLLHGEHLFKEYWNNPGATAEALADGWFTG DIGTLD
 DGYLRITGRKKEIIVTAGGKNVAPAVMEDRIRAHALVAECMVVGDGRPFV GALVTIDE
 EFLGRWCAEHGKPAGSTAVSLREDPELLAAIQDAVDDGNA AVSKAESVRKFRVLGAQF
 TEDSGHLTPSLKLRNVVAKDYADEIEAIYSK"

misc_feature 2050. .3387
 /gene="SC6G10.04"
 /note="Pfam match to entry PF00501 AMP-binding,
 AMP-binding enzyme, score -28.50, E-value 3.2e-13."

misc_feature 2458. .2493
 /gene="SC6G10.04"
 /note="PS00455 Putative AMP-binding domain signature."
 gene complement(3696. .4934)
 /gene="SC6G10.05c"
 CDS complement(3696. .4934)
 /gene="SC6G10.05c"
 /note="SC6G10.05c, possible glycosyl transferase, len: 412aa; similar to TR:053522 (EMBL:AL021957) hypothetical protein from Mycobacterium tuberculosis (399 aa) fasta scores; opt: 1276, z-score: 1452.2, E(): 0, (51.9% identity in 391 aa overlap) and SW:WCAL_SALTY putative colanic acid biosynthesis glycosyl transferase from the rfb (O antigen) gene cluster of Salmonella typhimurium (406 aa) fasta scores; opt: 291, z-score: 334.2, E(): 2.6e-11, (29.0% identity in 300 aa overlap). Contains Pfam match to entry PF00534 Glycos_transf_1, Glycosyl transferases group 1."
 /codon_start=1
 /transl_table=11
 /label=SC6G10.05c
 /product="putative glycosyl transferase"
 /protein_id="CAB39859.1"
 /db_xref="GI:4539201"
 /translation="MRKTLIVTNDFFPRPGGIQAFLHNMALRLDPERLVVYASTWKRG REGIEATAAFDAEQPFTVVRDRTTMLLPATPGATRRVGLLREHGCTSVWFGAAAPLGL MAPALRRAGAERLVATTHGHEAGWAQLPAARQLLRIGESTDTITYLGEYTRSRIAGALTPGAAARMVQLPPGVDEKTFHPASGGDEVRLRGFTDRPVVVCVSRVLPVPRKGQDTLI RAMPRILAAEPDAVLLIVGGGPYEKDLRLAEETGVAAAVHFTGAVPWSELPAHYGAG DVFAMPCRTTRGGGLDVEGLGIVYLEASATGLPVVAGDSGGAPDAVL DGETGWVVRGED PNESADRITLLADPELRRRMGERGRAWVEEKWRWDLLAEHLRLLQGGSAARARQAT DNVGPPTNRTRHPGRRPYLE"
 misc_feature complement(3849. .4322)
 /gene="SC6G10.05c"
 /note="Pfam match to entry PF00534 Glycos_transf_1, Glycosyl transferases group 1, score 86.70, E-value 1.1e-23."
 gene 5091. .6356
 /gene="SC6G10.06"
 CDS 5091. .6356
 /gene="SC6G10.06"

Query Match 9.6%; Score 39.6; DB 1; Length 36734;
 Best Local Similarity 48.6%; Pred. No. 3.9;
 Matches 108; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

Qy 162 ccgtctcatgtggcgtagcgtagcgcccgcgcatggttagagccccacccttcgctcgcaa 221
 || |||| | | ||| || ||| || | |||| |||| | | | |||
 Db 21826 CCCGTTTCATCGAGTCCTGGGTACCGGCGACAAGCGCGAGCACCATCCTGGACCGCCC 21767
 Qy 222 tcccatcaccatgaccctcacgcctggcgcgccgacacctccaagaaagtcgtgaa 281
 | || | || |||| | |||| | | |||| || || | ||||
 Db 21766 GCGCAACGCCCCGACCGTACGGCCTTCGGTGTGCCTGGCTGACCGTCTACTTCGTGCT 21707
 Qy 282 gacaagcactgtcttcttcccttctatgcaggtatccttggatggccagtcgcagccgc 341
 | | | || | | | | | | | | | | | | | | | | |
 Db 21706 GCTGATCGGTGGCGGCAACGACCTGTGGGCCACCACTTCCACCTGTGATCAACGCGAT 21647

Qy 342 ctggtggttcaacggaacatgtgactcttccaaatggaagt 383
| | | | | | | | | | | | | | | | | | | |
Db 21646 CACCTGGTTCGTCCGCATCGCGTTCTTCGTCGGACCGGTCGT 21605

RESULT 13

AC023212/c

LOCUS AC023212 78220 bp DNA HTG 13-JUL-2000

DEFINITION Homo sapiens clone RP11-758K16, LOW-PASS SEQUENCE SAMPLING.

ACCESSION AC023212

VERSION AC023212.2 GI:9164272

KEYWORDS HTG; HTGS_PHASE0.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 78220)

AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE Homo sapiens, clone RP11-758K16

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 78220)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczký,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (09-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT On Jul 13, 2000 this sequence version replaced gi:6957758.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L6564

Center clone name: 758_K_16

* NOTE: This record contains 89 individual

* sequencing reads that have not been assembled into

* contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

```
*      1      765: contig of 765 bp in length
*      766 865: gap of      100 bp
*      866     1651: contig of 786 bp in length
*     1652 1751: gap of      100 bp
*     1752     2513: contig of 762 bp in length
*     2514 2613: gap of      100 bp
*     2614     3394: contig of 781 bp in length
*     3395 3494: gap of      100 bp
*     3495     4270: contig of 776 bp in length
*     4271 4370: gap of      100 bp
*     4371     5160: contig of 790 bp in length
*     5161 5260: gap of      100 bp
*     5261     6049: contig of 789 bp in length
*     6050 6149: gap of      100 bp
*     6150     6933: contig of 784 bp in length
*     6934 7033: gap of      100 bp
*     7034     7801: contig of 768 bp in length
*     7802 7901: gap of      100 bp
*     7902     8736: contig of 835 bp in length
*     8737 8836: gap of      100 bp
*     8837     9629: contig of 793 bp in length
*     9630 9729: gap of      100 bp
*     9730    10507: contig of 778 bp in length
*    10508 10607: gap of      100 bp
*    10608    11385: contig of 778 bp in length
*    11386 11485: gap of      100 bp
*    11486    12268: contig of 783 bp in length
*    12269 12368: gap of      100 bp
*    12369    13151: contig of 783 bp in length
*    13152 13251: gap of      100 bp
*    13252    14017: contig of 766 bp in length
*    14018 14117: gap of      100 bp
*    14118    14905: contig of 788 bp in length
*    14906 15005: gap of      100 bp
*    15006    15786: contig of 781 bp in length
*    15787 15886: gap of      100 bp
*    15887    16664: contig of 778 bp in length
*    16665 16764: gap of      100 bp
*    16765    17541: contig of 777 bp in length
*    17542 17641: gap of      100 bp
*    17642    18383: contig of 742 bp in length
*    18384 18483: gap of      100 bp
*    18484    19266: contig of 783 bp in length
*    19267 19366: gap of      100 bp
*    19367    20128: contig of 762 bp in length
*    20129 20228: gap of      100 bp
*    20229    21009: contig of 781 bp in length
*    21010 21109: gap of      100 bp
```

* 21110 21881: contig of 772 bp in length
 * 21882 21981: gap of 100 bp
 * 21982 22831: contig of 850 bp in length
 * 22832 22931: gap of 100 bp
 * 22932 23699: contig of 768 bp in length
 * 23700 23799: gap of 100 bp
 * 23800 24570: contig of 771 bp in length
 * 24571 24670: gap of 100 bp
 * 24671 25428: contig of 758 bp in length
 * 25429 25528: gap of 100 bp
 * 25529 26301: contig of 773 bp in length
 * 26302 26401: gap of 100 bp
 * 26402 27167: contig of 766 bp in length
 * 27168 27267: gap of 100 bp
 * 27268 28048: contig of 781 bp in length
 * 28049 28148: gap of 100 bp
 * 28149 28920: contig of 772 bp in length
 * 28921 29020: gap of 100 bp
 * 29021 29796: contig of 776 bp in length
 * 29797 29896: gap of 100 bp
 * 29897 30687: contig of 791 bp in length
 * 30688 30787: gap of 100 bp
 * 30788 31565: contig of 778 bp in length
 * 31566 31665: gap of 100 bp
 * 31666 32431: contig of 766 bp in length
 * 32432 32531: gap of 100 bp
 * 32532 33298: contig of 767 bp in length
 * 33299 33398: gap of 100 bp
 * 33399 34168: contig of 770 bp in length
 * 34169 34268: gap of 100 bp
 * 34269 35051: contig of 783 bp in length
 * 35052 35151: gap of 100 bp
 * 35152 35921: contig of 770 bp in length
 * 35922 36021: gap of 100 bp
 * 36022 36766: contig of 745 bp in length
 * 36767 36866: gap of 100 bp
 * 36867 37642: contig of 776 bp in length
 * 37643 37742: gap of 100 bp
 * 37743 38517: contig of 775 bp in length
 * 38518 38617: gap of 100 bp
 * 38618 39384: contig of 767 bp in length
 * 39385 39484: gap of 100 bp
 * 39485 40262: contig of 778 bp in length
 * 40263 40362: gap of 100 bp
 * 40363 41140: contig of 778 bp in length
 * 41141 41240: gap of 100 bp
 * 41241 42017: contig of 777 bp in length
 * 42018 42117: gap of 100 bp
 * 42118 42859: contig of 742 bp in length
 * 42860 42959: gap of 100 bp
 * 42960 43743: contig of 784 bp in length
 * 43744 43843: gap of 100 bp
 * 43844 44628: contig of 785 bp in length
 * 44629 44728: gap of 100 bp
 * 44729 45506: contig of 778 bp in length
 * 45507 45606: gap of 100 bp
 * 45607 46386: contig of 780 bp in length

Query Match 9.6%; Score 39.6; DB 2; Length 78220;
Best Local Similarity 50.0%; Pred. No. 3.8;
Matches 93; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

Qy	82	accaatacaccccatcccaacactccaaacccaacaaacatttcaaccaaaccaccacaaca	141
Db	68204	ACCCACCCCCCCCCCCCCCCCCCCCCCCCCCAACACACCACCCCCACCCCCCAACA	68145
Qy	142	atgccttcagtaacccaggcccggtctcatgtggcgtagcgtcgcccgcggcatggtagag	201
Db	68144	AAAAAAACAAACCACNCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCC	68085
Qy	202	ccccaccccttcgctcgcaatcccatcaccatgacccctcaegcctggcgcgccgccgac	261
Db	68084	CCCCCCCCCCCCCCCCACCNCCCCCCCCCCCCACCNCCCCCCCCCCCCCCCCCCCCCCC	68025
Qy	262	ctctcc	267
Db	68024	CCCCCC	68019

RESULT 14

AC080179/c

LOCUS AC080179 54450 bp DNA HTG 16-JAN-2001

DEFINITION Homo sapiens chromosome X clone RP11-266P16 map X, LOW-PASS

SEQUENCE SAMPLING.

ACCESSION AC080179

VERSION AC080179.2 GI:12232512

KEYWORDS HTG; HTGS_PHASE0.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 54450)

AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE Homo sapiens chromosome X, clone RP11-266P16

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 54450)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K., Lamazares,R., Landers,T., Lehoczký,J., Levine,R., Lieu,C., Liu,G., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (28-SEP-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT On Jan 16, 2001 this sequence version replaced gi:10334899.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L11199

Center clone name: 266_P_16

* NOTE: This record contains 67 individual

* sequencing reads that have not been assembled into

* contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

```
*      1      684: contig of 684 bp in length
*      685 784: gap of      100 bp
*      785     1513: contig of 729 bp in length
*     1514 1613: gap of      100 bp
*     1614     2356: contig of 743 bp in length
*     2357 2456: gap of      100 bp
*     2457     3157: contig of 701 bp in length
*     3158 3257: gap of      100 bp
*     3258     3938: contig of 681 bp in length
*     3939 4038: gap of      100 bp
*     4039     4772: contig of 734 bp in length
*     4773 4872: gap of      100 bp
*     4873     5550: contig of 678 bp in length
*     5551 5650: gap of      100 bp
*     5651     6344: contig of 694 bp in length
*     6345 6444: gap of      100 bp
*     6445     7181: contig of 737 bp in length
*     7182 7281: gap of      100 bp
*     7282     8008: contig of 727 bp in length
*     8009 8108: gap of      100 bp
*     8109     8851: contig of 743 bp in length
*     8852 8951: gap of      100 bp
*     8952     9661: contig of 710 bp in length
*     9662 9761: gap of      100 bp
*     9762    10490: contig of 729 bp in length
*    10491 10590: gap of      100 bp
*    10591    11296: contig of 706 bp in length
*    11297 11396: gap of      100 bp
*    11397    12135: contig of 739 bp in length
*    12136 12235: gap of      100 bp
*    12236    12972: contig of 737 bp in length
*    12973 13072: gap of      100 bp
*    13073    13794: contig of 722 bp in length
*    13795 13894: gap of      100 bp
*    13895    14597: contig of 703 bp in length
*    14598 14697: gap of      100 bp
*    14698    15400: contig of 703 bp in length
*    15401 15500: gap of      100 bp
*    15501    16156: contig of 656 bp in length
*    16157 16256: gap of      100 bp
*    16257    16968: contig of 712 bp in length
*    16969 17068: gap of      100 bp
*    17069    17793: contig of 725 bp in length
*    17794 17893: gap of      100 bp
*    17894    18627: contig of 734 bp in length
*    18628 18727: gap of      100 bp
*    18728    19436: contig of 709 bp in length
*    19437 19536: gap of      100 bp
```


* 19537 20245: contig of 709 bp in length
 * 20246 20345: gap of 100 bp
 * 20346 21065: contig of 720 bp in length
 * 21066 21165: gap of 100 bp
 * 21166 21887: contig of 722 bp in length
 * 21888 21987: gap of 100 bp
 * 21988 22715: contig of 728 bp in length
 * 22716 22815: gap of 100 bp
 * 22816 23523: contig of 708 bp in length
 * 23524 23623: gap of 100 bp
 * 23624 24350: contig of 727 bp in length
 * 24351 24450: gap of 100 bp
 * 24451 25175: contig of 725 bp in length
 * 25176 25275: gap of 100 bp
 * 25276 25993: contig of 718 bp in length
 * 25994 26093: gap of 100 bp
 * 26094 26820: contig of 727 bp in length
 * 26821 26920: gap of 100 bp
 * 26921 27627: contig of 707 bp in length
 * 27628 27727: gap of 100 bp
 * 27728 28422: contig of 695 bp in length
 * 28423 28522: gap of 100 bp
 * 28523 29238: contig of 716 bp in length
 * 29239 29338: gap of 100 bp
 * 29339 30078: contig of 740 bp in length
 * 30079 30178: gap of 100 bp
 * 30179 30900: contig of 722 bp in length
 * 30901 31000: gap of 100 bp
 * 31001 31712: contig of 712 bp in length
 * 31713 31812: gap of 100 bp
 * 31813 32521: contig of 709 bp in length
 * 32522 32621: gap of 100 bp
 * 32622 33348: contig of 727 bp in length
 * 33349 33448: gap of 100 bp
 * 33449 34152: contig of 704 bp in length
 * 34153 34252: gap of 100 bp
 * 34253 34982: contig of 730 bp in length
 * 34983 35082: gap of 100 bp
 * 35083 35795: contig of 713 bp in length
 * 35796 35895: gap of 100 bp
 * 35896 36582: contig of 687 bp in length
 * 36583 36682: gap of 100 bp
 * 36683 37449: contig of 767 bp in length
 * 37450 37549: gap of 100 bp
 * 37550 38278: contig of 729 bp in length
 * 38279 38378: gap of 100 bp
 * 38379 39091: contig of 713 bp in length
 * 39092 39191: gap of 100 bp
 * 39192 39909: contig of 718 bp in length
 * 39910 40009: gap of 100 bp
 * 40010 40695: contig of 686 bp in length
 * 40696 40795: gap of 100 bp
 * 40796 41527: contig of 732 bp in length
 * 41528 41627: gap of 100 bp
 * 41628 42351: contig of 724 bp in length
 * 42352 42451: gap of 100 bp
 * 42452 43146: contig of 695 bp in length

* 43147 43246: gap of 100 bp
 * 43247 43972: contig of 726 bp in length
 * 43973 44072: gap of 100 bp
 * 44073 44791: contig of 719 bp in length
 * 44792 44891: gap of 100 bp
 * 44892 45608: contig of 717 bp in length
 * 45609 45708: gap of 100 bp
 * 45709 46431: contig of 723 bp in length
 * 46432 46531: gap of 100 bp
 * 46532 47231: contig of 700 bp in length
 * 47232 47331: gap of 100 bp
 * 47332 48036: contig of 705 bp in length
 * 48037 48136: gap of 100 bp
 * 48137 48852: contig of 716 bp in length
 * 48853 48952: gap of 100 bp
 * 48953 49694: contig of 742 bp in length
 * 49695 49794: gap of 100 bp
 * 49795 50475: contig of 681 bp in length
 * 50476 50575: gap of 100 bp
 * 50576 51276: contig of 701 bp in length
 * 51277 51376: gap of 100 bp
 * 51377 52115: contig of 739 bp in length
 * 52116 52215: gap of 100 bp
 * 52216 52898: contig of 683 bp in length
 * 52899 52998: gap of 100 bp
 * 52999 53731: contig of 733 bp in length
 * 53732 53831: gap of 100 bp
 * 53832 54450: contig of 619 bp in length.

FEATURES Location/Qualifiers
 source 1. .54450
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="X"

Query Match 9.5%; Score 39.2; DB 2; Length 54450;
 Best Local Similarity 44.1%; Pred. No. 4.9;
 Matches 86; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

Qy 73 cctagcgataccaatacacccatcccaacactccaaaccaaccaacacttcaaccaaacc 132
 || | | | | | | | | | | | | | | | | | | | | | |
 Db 47218 CCCANACACACAAAAAACAACCCACCCAACCAACCACACACAAAACCACCACCCC 47159
 Qy 133 accacaacaatgccttcagtaacccaggcccgctctcatgtggcgtagcgctcgcccgcggc 192
 || | | | | | | | | | | | | | | | | | | | | | |
 Db 47158 CCCCCNNCNCNNCCCCCNCCCACNNCCCCCCCCCCCCCCCCNNNNCCNCCCCCCCCCCCCCN 47099
 Qy 193 atggtagagccccaccccttcgctcgcaatcccatcaccatgacccctcacgcctggcgc 252
 || | | | | | | | | | | | | | | | | | | | | | |
 Db 47098 CCCNNCNCNNCCCCCCCCCCCCCNCCCCACCCNCCCCCCCCNCCCCNNCCNNNNNNNNCC 47039
 Qy 253 gccgcccgcctctcc 267
 || | | | | | | | | | | | | | | | | | | | | | |
 Db 47038 NCCCCCCNCCCCCC 47024

RESULT 15
 AE005747/c

LOCUS AE005747 11070 bp DNA BCT 28-MAR-2001
 DEFINITION *Caulobacter crescentus* section 73 of 359 of the complete genome.
 ACCESSION AE005747 AE005673
 VERSION AE005747.1 GI:13421943
 KEYWORDS .
 SOURCE *Caulobacter crescentus*.
 ORGANISM *Caulobacter crescentus*
 Bacteria; Proteobacteria; alpha subdivision; *Caulobacter* group;
Caulobacter.
 REFERENCE 1 (bases 1 to 11070)
 AUTHORS Nierman,W.C., Feldblyum,T.V., Laub,M.T., Paulsen,I.T., Nelson,K.E.,
 Eisen,J., Heidelberg,J.F., Alley,M.R.K., Ohta,N., Maddock,J.R.,
 Potocka,I., Nelson,W.C., Newton,A., Stephens,C., Phadke,N.D.,
 Ely,B., DeBoy,R.T., Dodson,R.J., Durkin,A.S., Gwinn,M.L.,
 Haft,D.H., Kolonay,J.F., Smit,J., Craven,M., Khouri,H., Shetty,J.,
 Berry,K., Utterback,T., Tran,K., Wolf,A., Vamathevan,J.,
 Ermolaeva,M., White,O., Salzberg,S.L., Venter,J.C., Shapiro,L. and
 Fraser,C.M.
 TITLE Complete genome sequence of *Caulobacter crescentus*
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4136-4141 (2001)
 MEDLINE 21173698
 REFERENCE 2 (bases 1 to 11070)
 AUTHORS Nierman,W.C., Feldblyum,T.V., Paulsen,I.T., Nelson,K.E., Eisen,J.,
 Heidelberg,J.F., Alley,M.R.K., Ohta,N., Maddock,J.R., Potocka,I.,
 Nelson,W.C., Newton,A., Stephens,C., Phadke,N.D., Ely,B.,
 Laub,M.T., DeBoy,R.T., Dodson,R.J., Durkin,A.S., Gwinn,M.L.,
 Haft,D.H., Kolonay,J.F., Smit,J., Craven,M., Khouri,H., Shetty,J.,
 Berry,K., Utterback,T., Tran,K., Wolf,A., Vamathevan,J.,
 Ermolaeva,M., White,O., Salzberg,S.L., Shapiro,L., Venter,J.C. and
 Fraser,C.M.
 TITLE Direct Submission
 JOURNAL Submitted (31-JAN-2001) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA
 FEATURES Location/Qualifiers
 source 1. .11070
 /organism="Caulobacter crescentus"
 /db_xref="taxon:69394"
 gene complement(110. .811)
 /gene="CC0717"
 CDS complement(110. .811)
 /gene="CC0717"
 /note="similar to GB:AE000513; identified by sequence
 similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="nodulin-related protein"
 /protein_id="AAK22702.1"
 /db_xref="GI:13421944"
 /translation="MTRLKVHVERHAVSRIGWLRAAVLGANDGIVSTASLVVGVA
 AAE ATRGPILLAAGAGLVAGAMSMAAGEYVSVASQADSEAADLARERAE
 LATQPEEELEEM TAIYVARGLTPDLARQVAEQLNAGDALAAHARDEL
 GISEHVTARPVQAALTSATFAV GAAMPLVVSL
 LAPLPVIIPTISVATLVFLAVLGWLGARTGGAS
 PWKPMLRVTFWGALA LLVTAVIGKLF
 GAVV"
 gene complement(959. .1435)
 /gene="CC0718"
 CDS complement(959. .1435)
 /gene="CC0718"

```

/feature "note" {
    /note="identified by Glimmer2; putative"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AAK22703.1"
    /db_xref="GI:13421945"
    /translation="MGSAPRAAQGAADVAGDERHSRAVAVIPPRDLHEHHREQLDALV
AACVLIAQVDGAVTPDERGRMVERLKLHAGLEGADLEAALRAFEALDARFDARPEETW
VEAELMIRRLKGSAAEAGVALAAVAVSIDGGLEAEERAAVLDICAWLGVSPPRVLP"
}
gene complement(1357..1626)
/feature "gene" {
    /gene="CC0719"
}
CDS complement(1357..1626)
/feature "note" {
    /note="identified by Glimmer2; putative"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AAK22704.1"
    /db_xref="GI:13421946"
    /translation="MLGPGGLPPLAIPPLARRRARPSTPEERPLTLLNARQLRAEIARL
TTALHEEARKPTPDRGLMRLWDLRRERLKEQLWSQETNATAARWR"
}
gene complement(1771..2829)
/feature "gene" {
    /gene="CC0720"
}
CDS complement(1771..2829)
/feature "note" {
    /note="identified by match to protein family HMM"
    /codon_start=1
    /transl_table=11
    /product="asparaginyl-tRNA synthetase, putative"
    /protein_id="AAK22705.1"
    /db_xref="GI:13421947"
    /translation="MSLPSSSPGAPWWRAGVHADRRPFLARNRITKAFAWFEVQGF
EEVEAAALQVSPGNEAHLHAFATQALTIAGEASPLYLHTSPEFACKKLLAAGENKIFS
LGKVWRNRERGPLHHPEFTMLEWYRAAPYQTLMDDCAVLLALAAETAGTQRFARFGM
TCDPFAAPERLSVAEAFQRHAGVDLLASVAADGSTDRDALTAQARAAGIRVAEDDGWA
DIFSRIIVEKVEPRLGEGRATILCEYPISEAALARPKDADPRVAERFELYACGVELAN
AFGELTDPAEQRRRFLIEMDEKARIYGERYPIDEDFLAALAHMPPASGSALGFDRVLV
LATGARRIEDVVWTPVAG"
}
gene 2950..3516
/feature "gene" {
    /gene="CC0721"
}
CDS 2950..3516
/feature "note" {
    /note="identified by match to TIGR protein family HMM
TIGR00038"
    /codon_start=1
    /transl_table=11
    /product="translation elongation factor P"
    /protein_id="AAK22706.1"
    /db_xref="GI:13421948"
    /translation="MKVAASSLRKGSVVDMDGKLYVVLSAENIHPGKGTPVTQLNMRR
ISDGVKVSERYRTTEQVERAFVDDRNHTFLYSDGDGYHFMNPESYDQLVATEDVIGDA
APYLQEGMTVILSTHNDVPIAIDLPRTVVLEIVDTEPSVKGQTASSSYKPAVLSNGVK
TTVPPYITAGTKVVILTEDGSYVERAKD"
}
gene complement(3716..6190)
/feature "gene" {
    /gene="CC0722"
}
CDS complement(3716..6190)
/feature "gene" {
    /gene="CC0722"
}

```

```

/feature "note" {
    note = "identified by match to protein family HMM"
}
/feature "codon_start" {
    codon_start = 1
}
/feature "transl_table" {
    transl_table = 11
}
/feature "product" {
    product = "TonB-dependent receptor"
}
/feature "protein_id" {
    protein_id = "AAK22707.1"
}
/feature "db_xref" {
    db_xref = "GI:13421949"
}
/feature "translation" {
    translation = "MRTKTKMTGHAALLCGAALCALLATAAQAETPATATPAASPD
DPNSVEQVIITSTRETRSAVAMEAQQIQRVLPGLKAIQFLPGVIYRTADPWGNNE
QNLSLFVHGFSTQQQLGYTMDGVPLGDQQYGNYNGLSVSRVTSENVSRVTLSSGAGSL
GVPSTSNLGGAIETFSRDPGAERSLTVNQTFGSYEASRTFLRAESGEMAGGFKGYLSY
LRQDAKAWDFDGRQRGHQVNLKLVREGEKGDLTIFYANWQMKVEPNEDATAMGNQQTAA
AQNFKPYSRAFIYDPDLGCTANLTGGPGTPPPAQGNFNSNCFSAAQREDILTYISYAW
KPTETLTWTNQAYYHYNFGRGIVSGPVNTAGLPGLFATYYPNLVVGTTTSTPTTLTNIV
NLFGGSGYAVRTTEYRINRPLISTVEWTLGDHDIKGGWFEHNEPAQQRAWYPMTKA
NNNLSPYDVPRGNKVLTYANAFFINNLQLHLQDQWRVTPKLLVQAGFKSSLQDATGK
FPVNQKNLPTVAVPVQFPTGSIENKWFPLPQAGLLWDATENDQVFVNQQNMRFIPY
GAGSGFYGFSAWSLGTQAQAFELFKKTVKPESTWYEAGLRSRREVAWGPIITGIEGQIN
AYNVKFSNRLLNIAPFNFINPAPAILANVGGVTTKGVDLAGTIKFGPRFQIYNAVSN
ESKYDSYQSGTTLVNGVSTPVTVAATGGKFVPLTPRWMNKTIARVNLGKFDAQISGDY
IGKRPVTYLNDLSVKAMMLVDLQAGYSFDIADGGMLKGLRLTANYTNLFDKKGVSTAG
VTSNSGGYVAFPPQPPRMGFVTVSAEF"
}
gene complement(6325..8100)
/feature "gene" {
    gene = "CC0723"
}
CDS complement(6325..8100)
/feature "gene" {
    gene = "CC0723"
}
/feature "note" {
    note = "identified by match to protein family HMM"
}
/feature "codon_start" {
    codon_start = 1
}
/feature "transl_table" {
    transl_table = 11
}
/feature "product" {
    product = "sensor histidine kinase / response regulator"
}
/feature "protein_id" {
    protein_id = "AAK22708.1"
}
/feature "db_xref" {
    db_xref = "GI:13421950"
}
/feature "translation" {
    translation = "MRPFATCKGAARTSLVSFPRTSAALPRFAEWLDQDVTSVTARS
LPTLGVRLAICAATALIYALAVDLQGGLLWGLAVASAEAAVFICSSPQRADRKISHAQ
RLSYVAVAWMNIVWWSLAIMLWRQDHPALMAALCVVCAFLVHAQAFTARSKTLLLI
VGGSATLILLVLGVLNDFPPSERLALCAAAILIIYAKAAQTNGQQGRALELAKSQ
AEAASQAKSEFLALMSHELRTPLNGLGLSQALKLEPLEPGEREQVELLEESGRLLA
LLNDVLDMAKIEAGKLDIAPTQEDLSRLAERVVRINQAQAREKGTEITLEIDPATPRA
LLFDPLRVQRCLGNLVSNAVKFTPAGQIRVRVTCEAGDQPDRLAKIIVSDTGVMGMP
AVLARIFTPEQADPTIARRTGGTGLGLNITRRLAQMMGGSVSVRSTEGKGSTFTLTF
GCRLPSAGAAETGGPLSGEHLRLLVDDYAVNRKVIAMMLAPLGCEIVEADNGQAL
DLAEREVDVLLDFNMPVMGGLETTRRLRADPRWRKLPVCLTAGVMDDERAAAATA
GMDGFIEKPIEMSTLVSTLARVARR"
}
gene complement(8170..8379)
/feature "gene" {
    gene = "CC0724"
}
CDS complement(8170..8379)
/feature "gene" {
    gene = "CC0724"
}
/feature "note" {
    note = "identified by Glimmer2; putative"
}
/feature "codon_start" {
    codon_start = 1
}
/feature "transl_table" {
    transl_table = 11
}
/feature "product" {
    product = "hypothetical protein"
}
/feature "protein_id" {
    protein_id = "AAK22709.1"
}
/feature "db_xref" {
    db_xref = "GI:13421951"
}
/feature "translation" {
    translation = "MIEELDTVALLIDKPALGLEKGLIGAVVYVHGAHEAFEVEFVDD
DGHTFCLETFKPEELLKIHRRAKAA"
}
gene complement(8376..8720)
/feature "gene" {
    gene = "CC0725"
}
CDS complement(8376..8720)

```

```

/gene="CC0725"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK22710.1"
/db_xref="GI:13421952"
/translation="MKLPGGESVQIDPRKLV DYALSPTHPVGKHKARLFASRLGLTQA
DAPALEAALRIAASQDATLIKSDAWGQRWQIDFMFQFGALSAMITSGWIVPADGSAV
RLTSVFIARDPG"
gene      complement(8725. .9666)
          /gene="CC0726"
CDS       complement(8725. .9666)
          /gene="CC0726"
          /note="identified by match to PFAM  protein family HMM
          PF00766"
          /codon_start=1
          /transl_table=11
          /product="electron transfer flavoprotein, alpha subunit"
          /protein_id="AAK22711.1"
          /db_xref="GI:13421953"
          /translation="MAVLVVADNDNALLRDATHKTVTAALKISGDVDVLVLGKGAKAV
ADQAAKIAGVRKVLLAESDALGHGVAEAQADAVLALAGNYDAILVPATSGGKNFAPRV
AAKLDVAPISEIVDVVAADTFTRPIYAGNALETVQSSDSKKVITVRPTAFAAAAEGGS
ASVESVSGADAAKTRFVSEEMVKSDRPELAAAKIVVSGGRAMGSAEEFQRVIEPLADK
LGAAVGASRAAVDAGYAPNDYQVGQTGKVVPQLYVAIGISGAIQHLAGMKDSKVIVA
INKDADAPIFQVADFLVADYKSAVPELMDALAAAGK"
gene      complement(9666. .10412)
          /gene="CC0727"

```

```

Query Match          9.5%;  Score 39;  DB 1;  Length 11070;
Best Local Similarity 51.4%;  Pred. No. 5.8;
Matches 90;  Conservative 0;  Mismatches 85;  Indels 0;  Gaps 0;

```

```

Qy   105 ccaaaccaaccaacacttcaaccaaaccaccacaacaatgccttcagtaacccaggcccg 164
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   2860 CGAAATCTCCAGCCTCTTCGGTTTTGCGGCCTTGTCCTGCCCTCCTCCTCCCCGGCGC 2801

Qy   165 tctcatgtggcgtagcgctcgcccgcgcatggttagagccccaccccttcgctcgcaatcc 224
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   2800 CCCATGGTGGCGCGCGCGGCGTCCACGCTGACCGCCGGCCGTTCTCCTGGCCCGCAACCG 2741

Qy   225 catcaccatgacccctcaagcctggcgcgccgcccacacctccaagaaagtcgtg 279
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   2740 GATCACCAAGGCGTTCCGCGCCTGGTTTCGAGGTTTCAGGGCTTCGAGGAAGTCGAG 2686

```

Search completed: February 7, 2002, 11:10:46
Job time: 10172 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 11:00:07 ; Search time 428.31 Seconds
 (without alignments)
 822.677 Million cell updates/sec

Title: US-09-394-745-6603
 Perfect score: 411
 Sequence: 1 agcaaaagcatagagatcca.....aggagaagaggaagggaccg 411

Scoring table: IDENTITY_NUC
 Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : N_Geneseq_1101:*
 1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
 2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
 3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
 4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
 5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
 6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
 7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
 8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
 9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
 10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
 11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
 12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
 13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
 14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
 15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
 16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
 17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
 18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
 19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
 20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
 21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
 22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

		%					
Result		Query					
No.	Score	Match	Length	DB	ID		Description
c	1	44.2	10.8	729	21	AAF09358	Fusarium venenatum
	2	39	9.5	4674	21	AAA14666	Nucleotide sequenc

	3	39	9.5	4725	21	AAA14665	Nucleotide sequenc
	4	39	9.5	4737	21	AAA14668	Nucleotide sequenc
	5	39	9.5	4767	21	AAA14667	Nucleotide sequenc
	6	39	9.5	4818	21	AAA14669	Nucleotide sequenc
	7	37	9.0	432	22	AAH52032	Mycobacterium tube
c	8	35.8	8.7	4932	20	AAX87629	Human insulin gene
	9	35.6	8.7	839	22	AAH00213	Bifidobacterium lo
	10	35.6	8.7	31422	21	AAA92302	S. avermitilis ave
	11	35.4	8.6	12381	21	AAZ58381	Streptomyces averm
	12	35.4	8.6	30690	21	AAA92301	S. avermitilis ave
c	13	34.8	8.5	114955	20	AAX53491	Human adenosine A1
	14	34.2	8.3	2634	19	AAV22334	Microbispora therm
	15	34	8.3	913	22	AAF67759	Corynebacterium gl
	16	34	8.3	1704	22	AAH65415	C glutamicum codin
c	17	34	8.3	349980	22	AAH68525	C glutamicum codin
	18	33.8	8.2	1973	22	AAI13556	Probe #3489 for ge
	19	33.8	8.2	1973	22	AAI34918	Probe #3604 used t
	20	33.8	8.2	1973	22	AAI03446	Probe #3437 used t
	21	33.4	8.1	1337	20	AAZ17263	Human gene express
	22	33.4	8.1	52253	21	AAA81478	N. meningitidis pa
c	23	33.4	8.1	349980	21	AAF21544	Neisseria meningit
c	24	33.4	8.1	1437668	21	AAA81490	N. meningitidis B
c	25	33.2	8.1	401	22	AAI14063	Probe #3996 for ge
c	26	33.2	8.1	401	22	AAI35443	Probe #4129 used t
c	27	33.2	8.1	401	22	AAI03921	Probe #3912 used t
c	28	33.2	8.1	446	22	AAI23276	Probe #13209 for g
c	29	33.2	8.1	446	22	AAI48596	Probe #17282 used
c	30	33.2	8.1	446	22	AAI08921	Probe #8912 used t
	31	33.2	8.1	936	22	AAF58252	Oligonucleotide D1
	32	33.2	8.1	936	22	AAF58254	Oligonucleotide D1
	33	33.2	8.1	936	22	AAF58257	Oligonucleotide D1
	34	33.2	8.1	936	22	AAF58259	Oligonucleotide D2
	35	33.2	8.1	936	22	AAF58262	Oligonucleotide D2
	36	33.2	8.1	938	22	AAF58255	Oligonucleotide D1
	37	33.2	8.1	1635	22	AAI22780	Probe #12713 for g
	38	33.2	8.1	1635	22	AAI48082	Probe #16768 used
	39	33.2	8.1	1635	22	AAI08454	Probe #8445 used t
	40	33	8.0	465	22	AAI10303	Probe #236 for gen
	41	33	8.0	465	22	AAI31554	Probe #240 used to
	42	33	8.0	465	22	AAI00239	Probe #230 used to
	43	33	8.0	822	22	AAH00987	Mycobacterium gord
	44	33	8.0	825	22	AAH00985	Mycobacterium aviu
	45	33	8.0	1827	22	AAF67757	Corynebacterium gl

ALIGNMENTS

RESULT 1
 AAF09358/c
 ID AAF09358 standard; cDNA; 729 BP.
 XX
 AC AAF09358;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Fusarium venenatum EST SEQ ID NO:1881.

XX
 KW Multiple gene expression; filamentous fungal cell; EST;
 KW expressed sequence tag; *Fusarium venenatum*; *Aspergillus niger*;
 KW *Aspergillus oryzae*; *Trichoderma reesei*; identification; recombination;
 KW culture condition; environmental stress; spore morphogenesis;
 KW metabolic pathway engineering; catabolic pathway engineering; ss.
 XX
 OS *Fusarium venenatum*.
 XX
 PN WO200056762-A2.
 XX
 PD 28-SEP-2000.
 XX
 PF 22-MAR-2000; 2000WO-US07781.
 XX
 PR 22-MAR-1999; 99US-0273623.
 XX
 PA (NOVO) NOVO NORDISK BIOTECH INC.
 PA (NOVO) NOVO NORDISK AS.
 XX
 PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
 XX
 DR WPI; 2000-594572/56.
 XX
 PT Monitoring differential expression of genes in filamentous fungal cells
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a
 PT substrate of expressed sequence tags -
 XX
 PS Claim 86; Page 1085-1086; 3161pp; English.
 XX
 CC The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring
 CC the global expression of genes from FF cells allows the production
 CC potential of the microorganisms to be improved. New genes may be
 CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how FF cells
 CC adapt to changes in culture conditions, environmental stress, spore
 CC morphogenesis, recombination, metabolic or catabolic pathway
 CC engineering. Using ESTs provides several advantages over genomic or
 CC random cDNA clones including elimination of redundancy as one spot on an
 CC array equals one gene or open reading frame, and organisation of the
 CC microarrays based on function of the gene products to facilitate
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
 CC *Fusarium venenatum*; AAF11248 to AAF11853 represents ESTs from *Aspergillus*
 CC *niger*; AAF11854 to AAF14878 represents ESTs from *Aspergillus oryzae*; and
 CC AAF14879 to AAF15337 represents ESTs from *Trichoderma reesei*, which are
 CC all specifically claimed in the present invention.
 XX
 SQ Sequence 729 BP; 202 A; 118 C; 202 G; 206 T; 1 other;

Query Match 10.8%; Score 44.2; DB 21; Length 729;
 Best Local Similarity 53.1%; Pred. No. 0.0011;
 Matches 94; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

```

Qy 184 gccgcggcatggttagagccccacccttcgctcgcaatcccatcaccatgaccctcac 243
    ||| || | | | | | | | | | | | | | | | | | | | | | | | |
Db 459 GCCGGCCGAGCTATGGAATCTCACCCCTTCGAGCGTCTTCCCCGAAGCTCAGAAGCCCGCT 400

Qy 244 gcctggcgcgccgcgcgacctctccaagaaagtcgtgaagacaagcactgtcttcttcccc 303
    | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 399 TCTCCTGATTACGCCAAGATGTTCAAGAGAGTTGGCAGCCAAGCCCTCTTCTTCTCCCT 340

Qy 304 ttctatgcaggtatccttggatggccagtcgcagccgcctggtggttcaacggaaac 360
    || || | | | | | | | | | | | | | | | | | | | | | |
Db 339 GGCTTCGCTGTCATCCTTGGCTGGCCTTTGGCTGCCAGTATGCCTTTGACGGTAGAC 283
  
```

RESULT 2

AAA14666

ID AAA14666 standard; DNA; 4674 BP.

XX

AC AAA14666;

XX

DT 08-AUG-2000 (first entry)

XX

DE Nucleotide sequence of modified FK-520 PKS gene cluster module 8.

XX

KW FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;
 KW immunophilin; FK-506 binding protein; polyketide compound; uveitis;
 KW transplant rejection; graft-versus-host disease; alopecia universalis;
 KW autoimmune chronic active hepatitis; inflammatory bowel disease;
 KW multiple sclerosis; primary biliary cirrhosis; scleroderma;
 KW neurite outgrowth; nerve regrowth; Parkinson's disease;
 KW Alzheimer's disease; stroke; traumatic spinal cord; brain injury;
 KW peripheral neuropathy; ss.

XX

OS Synthetic.

OS Streptomyces hygroscopicus.

XX

FH Key Location/Qualifiers

FT CDS 3..4673

FT /*tag= a

FT /note= "no termination codon given"

XX

PN WO200020601-A2.

XX

PD 13-APR-2000.

XX

PF 01-OCT-1999; 99WO-US22886.

XX

PR 02-OCT-1998; 98US-0102748.

PR 11-MAR-1999; 99US-0123810.

PR 17-JUN-1999; 99US-0139650.

XX

PA (KOSA-) KOSAN BIOSCIENCES INC.

XX

PI Reeves C, Chu D, Khosla C, Santi D, Wu K;
 XX
 DR WPI; 2000-317716/27.
 DR P-PSDB; AAY84730.
 XX
 PT New isolated polyketide synthase nucleic acid and polyketide compounds,
 PT useful for treating e.g. transplant rejection, uveitis, multiple
 PT sclerosis, Alzheimer's disease, Parkinson's disease, stroke, or
 PT peripheral neuropathy -
 XX
 PS Example 2; Page 93-96; 126pp; English.
 XX
 CC The present sequence represents module 8 of the FK-520 polyketide
 CC synthase (PKS) gene cluster, containing the acyltransferase (AT)
 CC domain of module 12 of rapamycin. FK-506 is a potent immunosuppressant,
 CC and acts through initial formation of an intermediate complex with
 CC protein immunophilins known as FK-506 binding proteins. The nucleic
 CC acids are used for producing polyketide compounds. The polyketide
 CC compounds can be used as immunosuppressants to prevent or treat
 CC transplant rejection, graft-versus-host disease or uveitis. They can
 CC also be used for treating e.g. alopecia universalis, autoimmune
 CC chronic active hepatitis, inflammatory bowel disease, multiple
 CC sclerosis, primary biliary cirrhosis, or scleroderma. They
 CC also have neurotrophic activity and can be used to promote neurite
 CC outgrowth in NGF-treated PC12 cells and in sensory neuronal cultures,
 CC and in intact animals, they promote regrowth of damaged facial and
 CC sciatic nerves, and repair lesioned serotonin and dopamine neurons in
 CC the brain. They can also be used for treating e.g. Parkinson's disease,
 CC Alzheimer's disease, stroke, traumatic spinal cord and brain injury, or
 CC peripheral neuropathies. They can also be used in agricultural and
 CC veterinary applications.
 XX
 SQ Sequence 4674 BP; 704 A; 1873 C; 1464 G; 633 T; 0 other;

Query Match 9.5%; Score 39; DB 21; Length 4674;
 Best Local Similarity 47.7%; Pred. No. 0.086;
 Matches 114; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

```

Qy 105 ccaaaccaaccaacacttcaaccaaaccaccacaacaatgccttcagtaacccaggcccg 164
      ||||| || |||| | ||| ||| | | | | | | | ||| ||
Db 4250 ccaaatcacccaagccctcacccacataccacaacccctcaccggcatcttccacaccgc 4309

Qy 165 tctcatgtggcgtagcgctcgcccgcgccatggttagagccccacccttcgctcgcaatcc 224
      || || || || | || | | |||| | || | || |
Db 4310 cgccaccctcgacgacgccaccctcaccaacctcacccccacacctcaccaccacct 4369

Qy 225 catcaccatgacccctcacgcctggcgcgccgcgacctctccaagaaagtcgtgaagac 284
      | ||| || ||||| || | | ||| || | || ||
Db 4370 ccaacccaaagccgacgcgcctggcacctccaccaccacacccaaaaccaacccctcac 4429

Qy 285 aagcactgtcttcttcccttctatgcaggatccttggatggccagtcgcagccgcct 343
      | ||| ||| || | || | || | ||| || | || | ||| ||
Db 4430 ccacttcgtcctctactccagcgccgcgcgccaccctcggcagcccgcccaagccaact 4488

```

RESULT 3

AAA14665

ID AAA14665 standard; DNA; 4725 BP.

XX

AC AAA14665;

XX

DT 08-AUG-2000 (first entry)

XX

DE Nucleotide sequence of FK-520 PKS gene cluster module 8.

XX

KW FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;
KW immunophilin; FK-506 binding protein; polyketide compound; uveitis;
KW transplant rejection; graft-versus-host disease; alopecia universalis;
KW autoimmune chronic active hepatitis; inflammatory bowel disease;
KW multiple sclerosis; primary biliary cirrhosis; scleroderma;
KW neurite outgrowth; nerve regrowth; Parkinson's disease;
KW Alzheimer's disease; stroke; traumatic spinal cord; brain injury;
KW peripheral neuropathy; ss.

XX

OS Streptomyces hygroscopicus.

XX

FH Key Location/Qualifiers

FT CDS 3..4724

FT /*tag= a

FT /note= "no termination codon given"

XX

PN WO200020601-A2.

XX

PD 13-APR-2000.

XX

PF 01-OCT-1999; 99WO-US22886.

XX

PR 02-OCT-1998; 98US-0102748.

PR 11-MAR-1999; 99US-0123810.

PR 17-JUN-1999; 99US-0139650.

XX

PA (KOSA-) KOSAN BIOSCIENCES INC.

XX

PI Reeves C, Chu D, Khosla C, Santi D, Wu K;

XX

DR WPI; 2000-317716/27.

DR P-PSDB; AAY84729.

XX

PT New isolated polyketide synthase nucleic acid and polyketide compounds,
PT useful for treating e.g. transplant rejection, uveitis, multiple
PT sclerosis, Alzheimer's disease, Parkinson's disease, stroke, or
PT peripheral neuropathy -

XX

PS Example 2; Page 90-93; 126pp; English.

XX

CC The present sequence encodes module 8 of the FK-520 polyketide
CC synthase (PKS) gene cluster of strain MA6548. FK-506 is a potent
CC immunosuppressants, and acts through initial formation of an
CC intermediate complex with protein immunophilins known as FK-506
CC binding proteins. The nucleic acids are used for producing polyketide
CC compounds. The polyketide compounds can be used as immunosuppressants to
CC prevent or treat transplant rejection, graft-versus-host disease or
CC uveitis. They can also be used for treating e.g. alopecia universalis,

CC autoimmune chronic active hepatitis, inflammatory bowel disease,
CC multiple sclerosis, primary biliary cirrhosis, or scleroderma. They
CC also have neurotrophic activity and can be used to promote neurite
CC outgrowth in NGF-treated PC12 cells and in sensory neuronal cultures,
CC and in intact animals, they promote regrowth of damaged facial and
CC sciatic nerves, and repair lesioned serotonin and dopamine neurons in
CC the brain. They can also be used for treating e.g. Parkinson's disease,
CC Alzheimer's disease, stroke, traumatic spinal cord and brain injury, or
CC peripheral neuropathies. They can also be used in agricultural and
CC veterinary applications.

XX

SQ Sequence 4725 BP; 728 A; 2034 C; 1394 G; 569 T; 0 other;

Query Match 9.5%; Score 39; DB 21; Length 4725;
Best Local Similarity 47.7%; Pred. No. 0.086;
Matches 114; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

```
Qy 105 ccaaaccaaccaacttcaacaaaccaccacaacaatgccttcagtaaccaggcccg 164
      ||||| || |||| | ||| ||| | | | | | | | | ||| ||
Db 4301 ccaaatcacccaagccctcacccacataccacaacccctcaccggcatcttcacaccgc 4360

Qy 165 tctcatgtggcgtagcgctcgcccgcgccatggtagagccccaccccttcgctcgcaatcc 224
      || || ||||| | || | ||||| | | || | || |
Db 4361 cgccaccctcgacgacgccaccctcaccaacctcaccacccaacacctcaccaccacct 4420

Qy 225 catcaccatgaccctcacgcctggcgcgccgacacctctccaagaaagtcgtgaagac 284
      | ||| || ||||| | | | ||| | || | ||
Db 4421 ccaacccaaagccgacgccgctggcacctccaccaccacacccaaaaccaaccctcac 4480

Qy 285 aagcactgtcttcttcccttctatgcaggtatccttggatggccagtcgcagccgcct 343
      | ||| ||| | || | || | ||| || | || | ||| ||
Db 4481 ccacttcgtcctctactccagcgccgcccaccctcggcagccccggccaagccaact 4539
```

RESULT 4

AAA14668

ID AAA14668 standard; DNA; 4737 BP.

XX

AC AAA14668;

XX

DT 08-AUG-2000 (first entry)

XX

DE Nucleotide sequence of modified FK-520 PKS gene cluster module 8.

XX

KW FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;
KW immunophilin; FK-506 binding protein; polyketide compound; uveitis;
KW transplant rejection; graft-versus-host disease; alopecia universalis;
KW autoimmune chronic active hepatitis; inflammatory bowel disease;
KW multiple sclerosis; primary biliary cirrhosis; scleroderma;
KW neurite outgrowth; nerve regrowth; Parkinson's disease;
KW Alzheimer's disease; stroke; traumatic spinal cord; brain injury;
KW peripheral neuropathy; ss.

XX

OS Synthetic.

OS Streptomyces hygroscopicus.

XX

FH Key Location/Qualifiers
 FT CDS 3..4736
 FT /*tag= a
 FT /note= "no termination codon given"
 XX
 PN WO200020601-A2.
 XX
 PD 13-APR-2000.
 XX
 PF 01-OCT-1999; 99WO-US22886.
 XX
 PR 02-OCT-1998; 98US-0102748.
 PR 11-MAR-1999; 99US-0123810.
 PR 17-JUN-1999; 99US-0139650.
 XX
 PA (KOSA-) KOSAN BIOSCIENCES INC.
 XX
 PI Reeves C, Chu D, Khosla C, Santi D, Wu K;
 XX
 DR WPI; 2000-317716/27.
 DR P-PSDB; AAY84732.
 XX
 PT New isolated polyketide synthase nucleic acid and polyketide compounds,
 PT useful for treating e.g. transplant rejection, uveitis, multiple
 PT sclerosis, Alzheimer's disease, Parkinson's disease, stroke, or
 PT peripheral neuropathy -
 XX
 PS Example 2; Page 99-102; 126pp; English.
 XX
 CC The present sequence represents module 8 of the FK-520 polyketide
 CC synthase (PKS) gene cluster, containing the acyltransferase (AT)
 CC domain of module 12 of rapamycin. FK-506 is a potent immunosuppressant,
 CC and acts through initial formation of an intermediate complex with
 CC protein immunophilins known as FK-506 binding proteins. The nucleic
 CC acids are used for producing polyketide compounds. The polyketide
 CC compounds can be used as immunosuppressants to prevent or treat
 CC transplant rejection, graft-versus-host disease or uveitis. They can
 CC also be used for treating e.g. alopecia universalis, autoimmune
 CC chronic active hepatitis, inflammatory bowel disease, multiple
 CC sclerosis, primary biliary cirrhosis, or scleroderma. They
 CC also have neurotrophic activity and can be used to promote neurite
 CC outgrowth in NGF-treated PC12 cells and in sensory neuronal cultures,
 CC and in intact animals, they promote regrowth of damaged facial and
 CC sciatic nerves, and repair lesioned serotonin and dopamine neurons in
 CC the brain. They can also be used for treating e.g. Parkinson's disease,
 CC Alzheimer's disease, stroke, traumatic spinal cord and brain injury, or
 CC peripheral neuropathies. They can also be used in agricultural and
 CC veterinary applications.
 XX
 SQ Sequence 4737 BP; 718 A; 1927 C; 1472 G; 620 T; 0 other;

Query Match 9.5%; Score 39; DB 21; Length 4737;
 Best Local Similarity 47.7%; Pred. No. 0.087;
 Matches 114; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

Qy 105 ccaaaccaaccaacacttcaaccaaaccaccacaacaatgccttcagtaacccaggcccg 164

```

      ||||| || |||| | ||| ||| | | | | | | | ||| ||
Db   4313 ccaaatacaccgaagccctcaccacataaccacaaccctcaccggcatcttcacaccgc 4372

Qy   165 tctcatgtggcgtagcgctcgcccgcgccatggtagagccccacccttcgctcgcaatcc 224
      ||          || | ||| | || | | ||||| | | ||| | || |
Db   4373 cgccaccctcgacgacgccaccctcaccaacctcaccaccaaacacctcaccaccaccct 4432

Qy   225 catcaccatgaccctcagcgctggcgcgccgacgtctccaagaaagtcgtgaagac 284
      |   |||  ||  ||||| || | | | ||| | || | ||
Db   4433 ccaacccaaagccgacgcccgtggcacctccaccaccacacccaaaaccaaccctcac 4492

Qy   285 aagcaactgtcttcttcccttctatgcaggtatccttggatggccagtcgcagccgcct 343
      |   ||| ||| | || | || | ||| || | || | ||| ||
Db   4493 ccacttcgtcctctactccagcgccgcccaccctcggcagccccggccaagccaact 4551

```

RESULT 5

AAA14667

ID AAA14667 standard; DNA; 4767 BP.

XX

AC AAA14667;

XX

DT 08-AUG-2000 (first entry)

XX

DE Nucleotide sequence of modified FK-520 PKS gene cluster module 8.

XX

KW FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;
KW immunophilin; FK-506 binding protein; polyketide compound; uveitis;
KW transplant rejection; graft-versus-host disease; alopecia universalis;
KW autoimmune chronic active hepatitis; inflammatory bowel disease;
KW multiple sclerosis; primary biliary cirrhosis; scleroderma;
KW neurite outgrowth; nerve regrowth; Parkinson's disease;
KW Alzheimer's disease; stroke; traumatic spinal cord; brain injury;
KW peripheral neuropathy; ss.

XX

OS Synthetic.

OS Streptomyces hygroscopicus.

XX

FH Key Location/Qualifiers

FT CDS 3..4766

FT /*tag= a

FT /note= "no termination codon given"

XX

PN WO200020601-A2.

XX

PD 13-APR-2000.

XX

PF 01-OCT-1999; 99WO-US22886.

XX

PR 02-OCT-1998; 98US-0102748.

PR 11-MAR-1999; 99US-0123810.

PR 17-JUN-1999; 99US-0139650.

XX

PA (KOSA-) KOSAN BIOSCIENCES INC.

XX

PI Reeves C, Chu D, Khosla C, Santi D, Wu K;

XX

DR WPI; 2000-317716/27.
 DR P-PSDB; AAY84731.
 XX
 PT New isolated polyketide synthase nucleic acid and polyketide compounds,
 PT useful for treating e.g. transplant rejection, uveitis, multiple
 PT sclerosis, Alzheimer's disease, Parkinson's disease, stroke, or
 PT peripheral neuropathy -
 XX
 PS Example 2; Page 96-99; 126pp; English.
 XX
 CC The present sequence represents module 8 of the FK-520 polyketide
 CC synthase (PKS) gene cluster, containing the acyltransferase (AT)
 CC domain of module 13 of rapamycin. FK-506 is a potent immunosuppressant,
 CC and acts through initial formation of an intermediate complex with
 CC protein immunophilins known as FK-506 binding proteins. The nucleic
 CC acids are used for producing polyketide compounds. The polyketide
 CC compounds can be used as immunosuppressants to prevent or treat
 CC transplant rejection, graft-versus-host disease or uveitis. They can
 CC also be used for treating e.g. alopecia universalis, autoimmune
 CC chronic active hepatitis, inflammatory bowel disease, multiple
 CC sclerosis, primary biliary cirrhosis, or scleroderma. They
 CC also have neurotrophic activity and can be used to promote neurite
 CC outgrowth in NGF-treated PC12 cells and in sensory neuronal cultures,
 CC and in intact animals, they promote regrowth of damaged facial and
 CC sciatic nerves, and repair lesioned serotonin and dopamine neurons in
 CC the brain. They can also be used for treating e.g. Parkinson's disease,
 CC Alzheimer's disease, stroke, traumatic spinal cord and brain injury, or
 CC peripheral neuropathies. They can also be used in agricultural and
 CC veterinary applications.
 XX
 SQ Sequence 4767 BP; 731 A; 1945 C; 1468 G; 623 T; 0 other;

Query Match 9.5%; Score 39; DB 21; Length 4767;
 Best Local Similarity 47.7%; Pred. No. 0.087;
 Matches 114; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

```

Qy 105 ccaaaccaaccaacacttcaaccaaaccaccacaacaatgccttcagtaaccaggcccg 164
      ||||| || |||| | ||| ||| | | | | | | | | ||| ||
Db 4343 ccaaatcacccaagccctcacccacataccacaacccctcaccggcatcttcacaccgc 4402

Qy 165 tctcatgtggcgtagcgctcgcccgcgccatggttagagccccaccccttcgctcgcaatcc 224
      || || || ||| | || | |||| | | || | || |
Db 4403 cgccaccctcgacgagccaccctcaccaacctcaccccccaacacctcaccaccaccct 4462

Qy 225 catcaccatgaccctcagcgctggcgcgccgacacctctccaagaaagtcgtgaagac 284
      | ||| || ||||| || | | ||| | || | ||
Db 4463 ccaacccaaagccgacgcccgtggcacctcaccaccacacccaaaaccaaccctcac 4522

Qy 285 aagcactgtcttcttcccttctatgcaggtatccttggatggccagtcgcagccgcct 343
      | ||| ||| | || | || | ||| || | || | ||| ||
Db 4523 ccacttcgtcctctactccagcgccgcccaccctcggcagccccggccaagccaact 4581

```

RESULT 6
 AAA14669
 ID AAA14669 standard; DNA; 4818 BP.

XX
 AC AAA14669;
 XX
 DT 08-AUG-2000 (first entry)
 XX
 DE Nucleotide sequence of modified FK-520 PKS gene cluster module 8.
 XX
 KW FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;
 KW immunophilin; FK-506 binding protein; polyketide compound; uveitis;
 KW transplant rejection; graft-versus-host disease; alopecia universalis;
 KW autoimmune chronic active hepatitis; inflammatory bowel disease;
 KW multiple sclerosis; primary biliary cirrhosis; scleroderma;
 KW neurite outgrowth; nerve regrowth; Parkinson's disease;
 KW Alzheimer's disease; stroke; traumatic spinal cord; brain injury;
 KW peripheral neuropathy; ss.
 XX
 OS Synthetic.
 OS Streptomyces hygroscopicus.
 XX
 FH Key Location/Qualifiers
 FT CDS 3..4817
 FT /*tag= a
 FT /note= "no termination codon given"
 XX
 PN WO200020601-A2.
 XX
 PD 13-APR-2000.
 XX
 PF 01-OCT-1999; 99WO-US22886.
 XX
 PR 02-OCT-1998; 98US-0102748.
 PR 11-MAR-1999; 99US-0123810.
 PR 17-JUN-1999; 99US-0139650.
 XX
 PA (KOSA-) KOSAN BIOSCIENCES INC.
 XX
 PI Reeves C, Chu D, Khosla C, Santi D, Wu K;
 XX
 DR WPI; 2000-317716/27.
 DR P-PSDB; AAY84733.
 XX
 PT New isolated polyketide synthase nucleic acid and polyketide compounds,
 PT useful for treating e.g. transplant rejection, uveitis, multiple
 PT sclerosis, Alzheimer's disease, Parkinson's disease, stroke, or
 PT peripheral neuropathy -
 XX
 PS Example 2; Page 102-105; 126pp; English.
 XX
 CC The present sequence represents module 8 of the FK-520 polyketide
 CC synthase (PKS) gene cluster, containing the acyltransferase (AT)
 CC domain of module 13 of rapamycin. FK-506 is a potent immunosuppressant,
 CC and acts through initial formation of an intermediate complex with
 CC protein immunophilins known as FK-506 binding proteins. The nucleic
 CC acids are used for producing polyketide compounds. The polyketide
 CC compounds can be used as immunosuppressants to prevent or treat
 CC transplant rejection, graft-versus-host disease or uveitis. They can
 CC also be used for treating e.g. alopecia universalis, autoimmune

CC chronic active hepatitis, inflammatory bowel disease, multiple
CC sclerosis, primary biliary cirrhosis, or scleroderma. They
CC also have neurotrophic activity and can be used to promote neurite
CC outgrowth in NGF-treated PC12 cells and in sensory neuronal cultures,
CC and in intact animals, they promote regrowth of damaged facial and
CC sciatic nerves, and repair lesioned serotonin and dopamine neurons in
CC the brain. They can also be used for treating e.g. Parkinson's disease,
CC Alzheimer's disease, stroke, traumatic spinal cord and brain injury, or
CC peripheral neuropathies. They can also be used in agricultural and
CC veterinary applications.

XX

SQ Sequence 4818 BP; 742 A; 1982 C; 1476 G; 618 T; 0 other;

Query Match 9.5%; Score 39; DB 21; Length 4818;
Best Local Similarity 47.7%; Pred. No. 0.087;
Matches 114; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

```
Qy 105 ccaaaccaaccaacttcaaccaaccaccacaacaatgccttcagtaaccaggcccg 164
      ||||| || |||| | ||| ||| | | | | | | | | | ||| ||
Db 4394 ccaaatcacccaagccctcacccacataccacaaccctcaccggcatcttccacaccgc 4453

Qy 165 tctcatgtggcgtagcgctcgcccgcgccatggtagagccccacccttcgctcgcaatcc 224
      || || | ||| | || | | |||| | | || | || |
Db 4454 cgccaccctcgacgacgccaccctcaccaacctccccccaacacctcaccaccacct 4513

Qy 225 catcaccatgaccctcagcctggcgcgccgacacctccaagaaagtcgtgaagac 284
      | ||| || ||||| || | | ||| | || | || ||
Db 4514 ccaacccaaagccgacgcgcctggcacctccaccaccacacccaaaaccaaccctcac 4573

Qy 285 aagcactgtcttcttcccttctatgcaggtatccttgatggccagtcgcagccgcct 343
      | ||| ||| || | || | || | ||| || ||| || ||| ||
Db 4574 ccacttcgtcctctactccagcgccgcccaccctcggcagccccggccaagccaact 4632
```

RESULT 7

AAH52032

ID AAH52032 standard; DNA; 432 BP.

XX

AC AAH52032;

XX

DT 04-SEP-2001 (first entry)

XX

DE Mycobacterium tuberculosis potential drug target gene SEQ ID 86.

XX

KW Drug target; growth; organism viability; characterisation; ds.

XX

OS Mycobacterium tuberculosis.

XX

PN WO200135317-A1.

XX

PD 17-MAY-2001.

XX

PF 13-NOV-2000; 2000WO-US31152.

XX

PR 12-NOV-1999; 99US-0165086.

PR 12-NOV-1999; 99US-0165124.

PR 01-FEB-2000; 2000US-0179531.
 XX
 PA (REGC) UNIV CALIFORNIA.
 XX
 PI Eisenberg D, Rotstein SH, Marcotte EM;
 XX
 DR WPI; 2001-329193/34.
 DR P-PSDB; AAG81181.
 XX
 PT Identifying nucleotide or polypeptide sequence for use as drug target,
 PT involves providing algorithm that analyzes a functional relationship
 PT between nucleotide or polypeptide sequences, and comparing the
 PT sequences -
 XX
 PS Disclosure; Page 105; 207pp; English.
 XX
 CC This invention relates to a method for identifying a nucleotide or
 CC polypeptide sequence that may be a drug target, or essential for growth
 CC or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092
 CC represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium
 CC tuberculosis proteins which are potential drug targets. The DNA and
 CC protein sequences are used to illustrate the method of the invention. The
 CC method involves providing an unknown nucleotide or polypeptide sequences,
 CC and comparing it to a number of sequences along with at least one
 CC algorithm capable of analysing a functional relationship between
 CC nucleotide and polypeptide sequences. The method is useful for
 CC characterising the function of nucleic acids and polypeptides that may be
 CC useful as a target for a drug or essential for the growth or viability of
 CC an organism.
 XX
 SQ Sequence 432 BP; 94 A; 139 C; 130 G; 69 T; 0 other;

Query Match 9.0%; Score 37; DB 22; Length 432;
 Best Local Similarity 53.0%; Pred. No. 0.14;
 Matches 79; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

Qy 155 cccaggcccggtctcatgtggcgtagcgtcgcccgcgcatggttagagccccacccttcg 214
 || |||| | | |||| |||| || | | | | || || |
 Db 188 ccgaggcgagagcggttcctacgtaatctcgccgcggtaccgacgaacagcatcccgaca 247
 Qy 215 ctcgcaatcccatcaccatgacccctcacgcctggcgcgccgacacctccaagaaag 274
 || | |||| || | | || | || || || || || || ||
 Db 248 gtcaaggccgatcaccttgctgcccaccaccgcccgtacgcaagcctttccaaggact 307
 Qy 275 tcgtgaagacaagcactgtcttcttcccc 303
 ||| || || | || | || |
 Db 308 gtgtggtgatcggcgcggtcgactatctc 336

RESULT 8
 AAX87629/c
 ID AAX87629 standard; DNA; 4932 BP.
 XX
 AC AAX87629;
 XX
 DT 26-OCT-1999 (first entry)

XX
 DE Human insulin gene.
 XX
 KW Insulin; preproinsulin; PPINS; human; epitope; autoantigen;
 KW autoantibody; insulin-dependent diabetes mellitus; IDDM;
 KW immunoassay; diagnosis; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 2364..3482
 FT /*tag= a
 FT /note= "contains an intron, exonic nucleotides
 FT 2424-2610 and 3397-3539 are included in
 FT the claimed cDNA of Claim 8"
 FT intron 2551..3336
 FT /*tag= b
 XX
 PN EP940470-A2.
 XX
 PD 08-SEP-1999.
 XX
 PF 29-DEC-1998; 98EP-0660149.
 XX
 PR 29-JAN-1998; 98US-0015399.
 XX
 PA (WALL-) WALLAC OY.
 XX
 PI Hinkkanen A;
 XX
 DR WPI; 1999-481070/41.
 DR P-PSDB; AAY06608.
 XX
 PT New fusion protein, useful for diagnosing insulin-dependent diabetes
 PT mellitus
 XX
 PS Claim 8; Page 19-22; 27pp; English.
 XX
 CC This is the nucleotide sequence of the human insulin gene coding
 CC for preproinsulin (PPINS, see AAY06608). The invention relates to a
 CC fusion protein having epitopes of at least 2 of the autoantigens
 CC glutamate decarboxylase (GAD65, see AAY06607), islet cell antigen
 CC (IA2, see AAY06606) and PPINS, in which the epitopes are connected
 CC via a linker peptide. The invention also provides cDNA encoding
 CC the fusion protein, which includes nucleotides 2424-2610 and
 CC 3397-3539 of the present sequence, a vector and an Escherichia coli
 CC cell encompassing the cDNA. The fusion protein is used in an
 CC immunoassay for the simultaneous detection of autoantibodies
 CC related to insulin-dependent diabetes mellitus (IDDM). Up to 3
 CC autoantibodies may be detected at once using the immunoassay. The
 CC presence of autoantibodies against multiple autoantigens is rare
 CC but is a strong indication of the (imminent) onset of IDDM, whereas
 CC the presence of autoantibodies to just one of the autoantigens may
 CC occur in healthy individuals.
 XX
 SQ Sequence 4932 BP; 835 A; 1531 C; 1736 G; 830 T; 0 other;

Query Match 8.7%; Score 35.8; DB 20; Length 4932;
Best Local Similarity 49.2%; Pred. No. 0.81;
Matches 94; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

```
Qy      81 taccaatacacccatcccaaacactccaaaccaaccaacacttcaaccaaaccaccacaac 140
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    1565 TCCCCACACCCCTGTCCCCAGACCCCTGTCCCCACACCCCTGTCCCCACACCCCTGTCCC 1506

Qy     141 aatgccttcagtaaacccaggcccggtctcatgtggcgtagcgctcgcccgcgccatggtaga 200
      | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    1505 CAGACCCCTGTCCCCACACCCCTGTCCCCGGACCCCTGTCCCCACACCCCTGTCCCCAGA 1446

Qy     201 gccccacccttcgctcgcaatcccatcaccatgacccctcacgcctggcgcgccgccga 260
      | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    1445 CCCCTGTCCCCACACCCCTGTCCCCACACCCCTGTCCCCAGACCCCTGTCCCCACACCCC 1386

Qy     261 cctctccaaga 271
      | | | | |
Db    1385 TGTCCCCAGGA 1375
```

RESULT 9

AAH00213

ID AAH00213 standard; DNA; 839 BP.

XX

AC AAH00213;

XX

DT 24-JUL-2001 (first entry)

XX

DE Bifidobacterium longum nucleotide sequence SEQ ID NO:204.

XX

KW Species specific; genus specific; family specific; probe; detection;

KW identification; algal; archaeal; bacterial; fungal; parasitical;

KW microorganism; diagnosis; translation elongation factor Tu; toxin;

KW translation elongation factor G; RecA recombinase; resistance;

KW catalytic subunit of proton-translocating ATPase; antimicrobial;

KW vaccine; primer; ds.

XX

OS Bifidobacterium longum.

XX

PN WO200123604-A2.

XX

PD 05-APR-2001.

XX

PF 28-SEP-2000; 2000WO-CA01150.

XX

PR 28-SEP-1999; 99CA-2283458.

PR 19-MAY-2000; 2000CA-2307010.

XX

PA (INFE-) INFECTIO DIAGNOSTIC (IDI) INC.

XX

PI Bergeron MG, Boissinot M, Huletsky A, Menard C, Ouellette M;

PI Picard FJ, Roy PH;

XX

DR WPI; 2001-245006/25.

XX

PT Nucleic acid sequences are used to generate universal probes and
PT primers which can be used to identify and detect the presence of algal,
PT archaeal, bacterial, fungal and parasitical species in a test sample -
XX
PS Claim 24; Page 527; 1580pp; English.
XX
CC The present invention describes a method for generating a repertory of
CC nucleic acids of tuf, fus, atpD and/or recA genes from which probes
CC and/or primers are derived. The method comprises amplifying the nucleic
CC acids of determined algal, archaeal, bacterial, fungal and parasitical
CC species with a combination of defined primer pairs. The method can be
CC used for producing probes and/or primers for detecting one or more
CC related microorganisms e.g. algae, archaea, bacteria, fungi and
CC parasites, for universal detection and for specific and ubiquitous
CC detection and identification of an algal, archaeal, bacterial, fungal
CC and parasitical species, genus, family and group. A nucleic acid (I)
CC obtained using the method of the invention can be used for the universal
CC detection of any bacterium, fungus or parasite in a sample and for the
CC detection of at least one antimicrobial agent resistance gene or at
CC least one toxin gene. hexA nucleic acids are used for the specific and
CC ubiquitous detection and for identification of Streptococcus pneumoniae.
CC (I) can be used to design a therapeutic agent which is effective against
CC microorganisms. Microbial species or genus or family or phylum or group
CC which can be detected include Abiotrophia adiacens, Bordetella sp.,
CC Corynebacterium sp., Enterobacteriaceae group, Escherichia coli,
CC Mycobacteriaceae family, Pseudomonads group, Streptococcus sp.,
CC Neisseria gonorrhoeae and Staphylococcus sp.. Using DNA based tests
CC provides faster results than substrate specificity tests as results can
CC be determined in an hour and improved accuracy is also achieved.
CC AAH00010 to AAH002304 represent nucleotide sequences and primers/probes
CC which are given in the exemplification of the present invention.
XX
SQ Sequence 839 BP; 152 A; 291 C; 241 G; 155 T; 0 other;

Query Match 8.7%; Score 35.6; DB 22; Length 839;
Best Local Similarity 48.1%; Pred. No. 0.47;
Matches 101; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

```

Qy      98 caacactccaaaccaaccaacacttcaaccaaaccaccacaacaatgccttcagtaaccc 157
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      471 ccaccgtcacctccatcgagaccttcacaagaccatggacgcctgcgaggctggcgaca 530

Qy      158 aggcccggtctcatgtggcgtagcgtagcgccgcggcgatggttagagccccacccttcgctc 217
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      531 acaccggtctgcttctgctgtggtctcgccggtgacgatgtcgagcgtggccaggttgtgg 590

Qy      218 gcaatcccatcaccatgacccctcacgcctggcgcgccgacgtctccaagaaagtgcg 277
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      591 ccaagccggggtccgtcaccgccgacaccaagttcgagggcggaagtctacgtgctgacca 650

Qy      278 tgaagacaagcactgtcttcttcccttct 307
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      651 aggacgaaggcgccggtcactcgccgttct 680

```

RESULT 10

AAA92302

ID AAA92302 standard; DNA; 31422 BP.

XX

AC AAA92302;

XX

DT 10-JAN-2001 (first entry)

XX

DE S. avermitilis avermectin aglycon synthase DNA aveAII SEQ ID NO:2.

XX

KW Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;

KW multifunctional enzyme; polyketide; avermectin; veterinary drug;

KW agrochemical; ds.

XX

OS Streptomyces avermitilis.

XX

FH Key Location/Qualifiers

FT CDS 1..14646

FT /*tag= a

FT /note= "avermectin aglycon synthase protein"

FT CDS 14824..31422

FT /*tag= b

FT /note= "avermectin aglycon synthase protein"

XX

PN WO200050605-A1.

XX

PD 31-AUG-2000.

XX

PF 23-FEB-2000; 2000WO-JP01041.

XX

PR 24-FEB-1999; 99JP-0046961.

XX

PA (KITA) KITASATO INST.

XX

PI Omura S, Ikeda H;

XX

DR WPI; 2000-565458/52.

DR P-PSDB; AAB23751, AAB23752.

XX

PT Avermectin aglycone synthase DNA and proteins encoded by all or part of

PT it for the production of avermectin and its derivatives for drug and

PT agrochemical use -

XX

PS Claim 2; Page 134-203; 314pp; Japanese.

XX

CC The present sequence represents DNA which encodes avermectin aglycon
CC synthase proteins. Also described are: (1) polypeptides encoded by all
CC or part of the DNA; (2) expression vectors containing the DNA; (3) host
CC cells transformed by the vectors; (4) preparation of the polypeptides
CC by culture of the transformants; (5) preparation of avermectin aglycon
CC or its derivatives by culture of transformed avermectin-producing
CC microorganisms; and (6) oligonucleotides of 5-60 bases in length
CC containing sense or antisense sequences from the avermectin aglycon
CC synthase DNA. The enzymes are useful for the production of modified
CC forms of avermectin and of the intermediates in its biosynthesis, for
CC use as drugs, veterinary drugs and agrochemicals.

XX

SQ Sequence 31422 BP; 4136 A; 10238 C; 11677 G; 5371 T; 0 other;

Query Match 8.7%; Score 35.6; DB 21; Length 31422;
Best Local Similarity 49.5%; Pred. No. 1.9;
Matches 92; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

```
Qy      76 agcgataccaatacacccatcccaacactccaaaccaacacacttcaaccaaaccacc 135
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 28258 atcgaaatcagtcacccacccaccctcgctcccgccatcgaagacaccaccgaaaacacc 28317

Qy     136 acaacaatgccttcagtaaccagggccggtctcatgtggcgtagcgtcgcccgcggcatg 195
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 28318 accgaaaacatcaccgcgacccggcagcctccgcccggcgacaacgacacccaccgcttc 28377

Qy     196 gtagagccccaccccttcgctcgcaatcccatcaccatgaccctcacgcctggcgcgcc 255
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 28378 ctcaccgcctcgcccacacccacaccacggcattcggacacccaccacctggcaccac 28437

Qy     256 gccgac 261
      | |
Db 28438 cactac 28443
```

RESULT 11

AAZ58381

ID AAZ58381 standard; DNA; 12381 BP.

XX

AC AAZ58381;

XX

DT 23-MAY-2000 (first entry)

XX

DE Streptomyces avermitilis avermectin polyketide synthase modules 1+2.

XX

KW Polyketide synthase; avermectin; insecticide; ss.

XX

OS Streptomyces avermitilis.

XX

PN WO200001827-A2.

XX

PD 13-JAN-2000.

XX

PF 06-JUL-1999; 99WO-GB02158.

XX

PR 06-JUL-1998; 98GB-0014622.

XX

PA (BIOT-) BIOTICA TECHNOLOGY LTD.

PA (PFIZ) PFIZER INC.

XX

PI Kellenberger JL, Leadlay PF, Staunton J, Stutzman-Engwall KJ;

PI McArthur HAI;

XX

DR WPI; 2000-182117/16.

XX

PT Mutated Type I polyketide synthase containing a polylinker site in an

PT extension module for replacement of a reductive loop sequence, for

PT producing polyketides, e.g. B1 avermectin -

XX

PS Disclosure; Fig 7a-f; 75pp; English.
 XX
 CC The present sequence is that of DNA encoding the first 2 modules
 CC of the avermectin polyketide synthase (PKS) of Streptomyces
 CC avermitilis. The invention relates to nucleic acids encoding a
 CC Type I PKS such as avermectin in which a polylinker with multiple
 CC restriction sites replaces or 1 more PKS genes encoding enzymes
 CC associated with reduction. Novel PKS are provided in which in
 CC which the reductive loop in a selected module of the Type I PKS is
 CC replaced with the equivalent segment from the same or different
 CC PKS gene cluster or by a mutated or synthetic segment. Vectors and
 CC host cells, and methods for producing novel polyketides by
 CC culturing host cells are claimed. The polyketides obtained are
 CC useful as antibiotics and insecticides. Fermentation products
 CC containing C22-C23 dihydroavermectin, ivermectin and B1
 CC avermectins are claimed.
 XX
 SQ Sequence 12381 BP; 1884 A; 4561 C; 4005 G; 1931 T; 0 other;

Query Match 8.6%; Score 35.4; DB 21; Length 12381;
 Best Local Similarity 49.2%; Pred. No. 1.5;
 Matches 93; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

Qy 96 cccaacactccaaaccaacacacttcaaccaaaccaccacaacaatgccttcagtaac 155
 || || |||| |||| ||| || | ||||| | | |||
 Db 5308 octcaccctccccaccacccaccaacccaccaaactggctcatcgccatccccgaaac 5367
 Qy 156 ccaggcccggtctcatgtggcgtagcgtagcgccgcggcatggtagagccccacccttcgc 215
 |||| ||| | | | | | | | | | | | | | | |
 Db 5368 ccagaccaccacccccacatcaccaacatcctcaccaacctccaccaccacggcatcac 5427
 Qy 216 tcgcaatcccatcaccatgaccctcacgcctggcgcgccgacgtctccaagaaagt 275
 | || ||| |||| | || | | | | || |||| | | |
 Db 5428 cccatccccctcacctcaaccacacccacaccaacccccaacacctccaccacaccct 5487
 Qy 276 cgtgaagac 284
 | | ||
 Db 5488 ccaccacac 5496

RESULT 12
 AAA92301
 ID AAA92301 standard; DNA; 30690 BP.
 XX
 AC AAA92301;
 XX
 DT 10-JAN-2001 (first entry)
 XX
 DE S. avermitilis avermectin aglycon synthase DNA aveAI SEQ ID NO:1.
 XX
 KW Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
 KW multifunctional enzyme; polyketide; avermectin; veterinary drug;
 KW agrochemical; ds.
 XX
 OS Streptomyces avermitilis.
 XX

FH Key Location/Qualifiers
 FT CDS 1..11919
 FT /*tag= a
 FT /note= "avermectin aglycon synthase protein"
 FT CDS 11971..30690
 FT /*tag= b
 FT /note= "avermectin aglycon synthase protein"
 XX
 PN WO200050605-A1.
 XX
 PD 31-AUG-2000.
 XX
 PF 23-FEB-2000; 2000WO-JP01041.
 XX
 PR 24-FEB-1999; 99JP-0046961.
 XX
 PA (KITA) KITASATO INST.
 XX
 PI Omura S, Ikeda H;
 XX
 DR WPI; 2000-565458/52..
 DR P-PSDB; AAB23749, AAB23750.
 XX
 PT Avermectin aglycone synthase DNA and proteins encoded by all or part of
 PT it for the production of avermectin and its derivatives for drug and
 PT agrochemical use -
 XX
 PS Claim 2; Page 66-134; 314pp; Japanese.
 XX
 CC The present sequence represents DNA which encodes avermectin aglycon
 CC synthase proteins. Also described are: (1) polypeptides encoded by all
 CC or part of the DNA; (2) expression vectors containing the DNA; (3) host
 CC cells transformed by the vectors; (4) preparation of the polypeptides
 CC by culture of the transformants; (5) preparation of avermectin aglycon
 CC or its derivatives by culture of transformed avermectin-producing
 CC microorganisms; and (6) oligonucleotides of 5-60 bases in length
 CC containing sense or antisense sequences from the avermectin aglycon
 CC synthase DNA. The enzymes are useful for the production of modified
 CC forms of avermectin and of the intermediates in its biosynthesis, for
 CC use as drugs, veterinary drugs and agrochemicals.
 XX
 SQ Sequence 30690 BP; 5356 A; 12454 C; 8617 G; 4263 T; 0 other;

Query Match 8.6%; Score 35.4; DB 21; Length 30690;
 Best Local Similarity 49.2%; Pred. No. 2.2;
 Matches 93; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

Qy 96 cccaacactccaaaccaaccaacttcaaccaaaccaccacaacaatgccttcagtaac 155
 || || |||| |||| ||| || | ||||| | | |||
 Db 4491 cctcaccctccccaccaccaccaacccaccaaacctggctcatcgccatccccgaaac 4550
 Qy 156 ccaggcccggtctcatgtggcgtagcgtagcgccgcggcatggttagagccccacccttcgc 215
 |||| ||| | | | | | | | | | | | | | | | |
 Db 4551 ccagaccacacccccacatcaccaacatcctcaccaacctccaccaccacggcatcac 4610
 Qy 216 tcgcaatcccatcaccatgaccctcacgcctggcgcgccgacacctctccaagaaagt 275

Db 4611 ccccatccccctcaccctcaaccacacccacaccaaccccccaacacctccaccacaccct 4670
 Qy 276 cgtgaagac 284
 Db 4671 ccaccacac 4679

RESULT 13

AAX53491/c

ID AAX53491 standard; DNA; 114955 BP.

XX

AC AAX53491;

XX

DT 05-JUL-1999 (first entry)

XX

DE Human adenosine A1 receptor antisense oligonucleotide fragment.

XX

KW Antisense oligonucleotide; multiple target; antisense treatment;

KW impaired respiration; inflammation; lung disease;

KW pulmonary vasoconstriction; inflammation; allergic rhinitis;

KW acute asthma; allergy; asthma; impeded respiration;

KW respiratory distress syndrome; pain; cystic fibrosis;

KW pulmonary hypertension; pulmonary vasoconstriction; emphysema;

KW chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;

KW colon cancer; breast cancer; lung cancer; pancreatic cancer;

KW hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;

KW prostate cancer; ss.

XX

OS Synthetic.

XX

PN WO9913886-A1.

XX

PD 25-MAR-1999.

XX

PF 17-SEP-1998; 98WO-US19419.

XX

PR 09-JUN-1998; 98US-0093972.

PR 17-SEP-1997; 97US-0059160.

XX

PA (UYEC-) UNIV EAST CAROLINA.

XX

PI Nyce JW;

XX

DR WPI; 1999-229400/19.

XX

PT New antisense oligonucleotides used in treatment of, e.g. pulmonary
 PT vasoconstriction

XX

PS Disclosure; Page 37; 120pp; English.

XX

CC The specification describes antisense oligonucleotides (AAX52869-X55271)
 CC directed against at least 2 mRNAs selected from target genes, coding and
 CC non-coding regions of RNAs corresponding to target genes, gene
 CC initiation codons, genomic flanking regions, intron-exon borders, the
 CC 5'-end, the 3'-end and the juxta-section between coding and non-coding
 CC regions and all segments of RNAs encoding proteins associated with one

CC or more diseases, conditions or mixtures. The antisense oligonucleotides
 CC may be derived from sequences AAX55272-74. These multiple target
 CC oligonucleotides (specifically AAX55180-271) can be used for the
 CC antisense treatment of diseases and conditions. Typical diseases and
 CC conditions are those associated with impaired respiration and
 CC inflammation, including lung diseases, pulmonary vasoconstriction,
 CC inflammation, allergic rhinitis, acute asthma, allergies, asthma, impeded
 CC respiration, respiratory distress syndrome, pain, cystic fibrosis,
 CC pulmonary hypertension, pulmonary vasoconstriction, emphysema, chronic
 CC obstructive pulmonary disease (COPD), and cancers such as leukemias,
 CC lymphomas, carcinomas e.g. colon cancer, breast cancer, lung cancer,
 CC pancreatic cancer, hepatocellular carcinoma, kidney cancer, melanoma,
 CC hepatic metastases, as well as all types of cancers which may metastasize
 CC or have metastasized to the lungs, including breast and prostate cancer.
 XX
 SQ Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 21427 other;

Query Match 8.5%; Score 34.8; DB 20; Length 114955;
 Best Local Similarity 43.5%; Pred. No. 5.5;
 Matches 74; Conservative 7; Mismatches 89; Indels 0; Gaps 0;

Qy 173 ggcgtagcgctcgcccgcgccatggtagagccccaccccttcgctcgcaatcccatcacca 232
 |||| || | |||| : : |||| || | ||| | :
 Db 108267 GGCGCCGCCGCCCGCCNNHNNNSCCVAGGCGAGCCAGGCGCCGCCGCCCGCCNNHN 108208

Qy 233 tgaccacctcacgcctggcgcgccgacacctccaagaaagtcgtgaagacaagcactg 292
 |||: || ||||| || || : | : | | ||
 Db 108207 NNSCCCVAGGCGAGCCAGGCGCCGCCGCCCGCCNNHNNNSGCCCVAGGCGAGCCAGGC 108148

Qy 293 tcttcttccccttctatgcaggtatccttgatggccagtcgcagccgcc 342
 | | |||| | : || || |||| || ||||
 Db 108147 GCCGCCGCCCGCCNNHNNNSGGCCCVAGGCGAGCCAGGCGCCGCCGCC 108098

RESULT 14
 AAV22334
 ID AAV22334 standard; DNA; 2634 BP.
 XX
 AC AAV22334;
 XX
 DT 17-AUG-1998 (first entry)
 XX
 DE Microbispora thermorosea pyruvate orthophosphate dikinase gene.
 XX
 KW Pyruvate orthophosphate dikinase; PPDk; pyrophosphoric acid;
 KW assay; ds.
 XX
 OS Microbispora thermorosea strain IFO 14047.
 XX
 PN GB2317892-A.
 XX
 PD 08-APR-1998.
 XX
 PF 02-OCT-1997; 97GB-0021083.
 XX
 PR 03-OCT-1996; 96JP-0281304.

DE Corynebacterium glutamicum MCT protein encoding DNA SEQ ID NO:33.
 XX
 KW Corynebacterium glutamicum; brevibacterium lactofermentum; MCT;
 KW membrane construction and membrane transport protein; petroleum spill;
 KW hydrocarbon degradation; gram positive aerobic bacterium; marker;
 KW identification; microorganism; fine chemical production; transformation;
 KW genome mapping; genetic engineering; ds.
 XX
 OS Corynebacterium glutamicum.
 XX
 PN WO200100805-A2.
 XX
 PD 04-JAN-2001.
 XX
 PF 23-JUN-2000; 2000WO-IB00926.
 XX
 PR 25-JUN-1999; 99US-0141031.
 PR 08-JUL-1999; 99DE-1031454.
 PR 08-JUL-1999; 99DE-1031478.
 PR 08-JUL-1999; 99DE-1031563.
 PR 09-JUL-1999; 99DE-1032122.
 PR 09-JUL-1999; 99DE-1032124.
 PR 09-JUL-1999; 99DE-1032125.
 PR 09-JUL-1999; 99DE-1032128.
 PR 09-JUL-1999; 99DE-1032180.
 PR 09-JUL-1999; 99DE-1032182.
 PR 09-JUL-1999; 99DE-1032190.
 PR 09-JUL-1999; 99DE-1032191.
 PR 09-JUL-1999; 99DE-1032209.
 PR 09-JUL-1999; 99DE-1032212.
 PR 09-JUL-1999; 99DE-1032227.
 PR 09-JUL-1999; 99DE-1032228.
 PR 09-JUL-1999; 99DE-1032229.
 PR 09-JUL-1999; 99DE-1032230.
 PR 14-JUL-1999; 99DE-1032927.
 PR 14-JUL-1999; 99DE-1033005.
 PR 14-JUL-1999; 99DE-1033006.
 PR 27-AUG-1999; 99DE-1040764.
 PR 27-AUG-1999; 99DE-1040765.
 PR 27-AUG-1999; 99DE-1040766.
 PR 27-AUG-1999; 99DE-1040830.
 PR 27-AUG-1999; 99DE-1040831.
 PR 27-AUG-1999; 99DE-1040832.
 PR 27-AUG-1999; 99DE-1040833.
 PR 31-AUG-1999; 99DE-1041378.
 PR 31-AUG-1999; 99DE-1041379.
 PR 31-AUG-1999; 99DE-1041395.
 PR 03-SEP-1999; 99DE-1042077.
 PR 03-SEP-1999; 99DE-1042078.
 PR 03-SEP-1999; 99DE-1042079.
 PR 03-SEP-1999; 99DE-1042088.
 XX
 PA (BADI) BASF AG.
 XX
 PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
 XX
 DR WPI; 2001-071486/08.

DR P-PSDB; AAB76526.
 XX
 PT Corynebacterium glutamicum nucleic acids encoding membrane construction
 PT and membrane transport proteins or their portions, useful for typing or
 PT identifying C. glutamicum or related bacteria, and as markers for
 PT transformation -
 XX
 PS Claim 3; Page 185-186; 1119pp; English.
 XX
 CC AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane
 CC construction and membrane transport (MCT) proteins given in AAB76510 to
 CC AAB76847. The MCT nucleic acids and proteins are useful in the
 CC identification of microorganisms which can be used to produce fine
 CC chemicals, for modulating fine chemical production in C. glutamicum or
 CC related bacteria (e.g. Brevibacterium lactofermentum), the typing or
 CC identification of C. glutamicum or related bacteria, as reference points
 CC for mapping C. glutamicum genome, and as markers for transformation.
 CC AAF68082 and AAF68082 represent sequencing primers which are used in an
 CC example from the present invention.
 XX
 SQ Sequence 913 BP; 233 A; 321 C; 206 G; 153 T; 0 other;

Query Match 8.3%; Score 34; DB 22; Length 913;
 Best Local Similarity 48.9%; Pred. No. 1.5;
 Matches 91; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

Qy 118 cacttcaaccaaaccaccacaacaatgccttcagtaacccaggcccgtctcatgtggcgt 177
 | | | | | | | | | | | | | | | | | | | | | |
 Db 638 ctctccggcgcgagcccccaacacggttccttttgaaaccctgaccagcgagcaatgggc 697
 Qy 178 agcgtcgccccgcgcatggttagagccccacccttcgctcgcaatcccatcaccatgacc 237
 | | | | | | | | | | | | | | | | | | | | | |
 Db 698 ggcgacggcgacgacgtagtttcagaaccacggtagacaaagaatccgctcgcgctgatc 757
 Qy 238 cctcacgcctggcgcgccgacgtctccaagaaagtcgtgaagacaagcactgtcttc 297
 | | | | | | | | | | | | | | | | | | | | | |
 Db 758 ctctacacctccggcaccacgggacgcccccaagggtgccagctcaccacggaaacctg 817
 Qy 298 ttcccc 303
 | | | |
 Db 818 ttctcc 823

Search completed: February 7, 2002, 11:00:37
 Job time: 5023 sec

GenCore version 4.5
 Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 11:12:24 ; Search time 172.96 Seconds
 (without alignments)
 538.173 Million cell updates/sec

Title: US-09-394-745-6603
 Perfect score: 411
 Sequence: 1 agcaaaagcatagagatcca.....aggagaagaggaagggaccg 411

Scoring table: IDENTITY_NUC
 Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : Issued_Patents_NA:*
 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
 5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
 6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	%					Description
	No.	Score	Match	Length	ID	
c	1	35	8.5	289	4	US-09-007-005-17
c	2	35	8.5	289	4	US-09-244-796-17
	3	35	8.5	7218	1	US-08-232-463-14
	4	34.2	8.3	2634	3	US-08-941-936-1
	5	33	8.0	1280	4	US-09-060-756-4
	6	32.8	8.0	925	3	US-08-858-003-1
	7	32.8	8.0	925	3	US-09-078-166-1
	8	32.8	8.0	925	4	US-08-997-467-1
	9	32.6	7.9	49377	1	US-08-764-233A-1
c	10	31.4	7.6	490	4	US-08-998-416-882
c	11	31.2	7.6	30001	1	US-08-125-468-1
c	12	31.2	7.6	30001	2	US-08-474-933-1
	13	30.8	7.5	11495	4	US-09-056-105-9
	14	30.6	7.4	35081	2	US-08-752-760A-1
	15	30.6	7.4	4403765	4	US-09-103-840A-2
	16	30.4	7.4	696	4	US-09-378-255-7
	17	30.4	7.4	696	4	US-09-715-336-7
	18	30.4	7.4	2669	4	US-09-141-212-9
	19	30.4	7.4	2669	4	US-09-561-138-9
	20	30.4	7.4	3255	2	US-08-916-917-11
	21	30.4	7.4	3255	3	US-09-225-170-11

22	30.4	7.4	3630	4	US-09-378-255-5	Sequence 5, Appli
23	30.4	7.4	3630	4	US-09-141-212-1	Sequence 1, Appli
24	30.4	7.4	3630	4	US-09-141-212-3	Sequence 3, Appli
25	30.4	7.4	3630	4	US-09-552-351-3	Sequence 3, Appli
26	30.4	7.4	3630	4	US-09-251-372-5	Sequence 5, Appli
27	30.4	7.4	3630	4	US-09-561-138-1	Sequence 1, Appli
28	30.4	7.4	3630	4	US-09-561-138-3	Sequence 3, Appli
29	30.4	7.4	3630	4	US-09-715-336-5	Sequence 5, Appli
30	30.2	7.3	152331	3	US-09-128-155-16	Sequence 16, Appl
31	30	7.3	714	4	US-09-105-537-17	Sequence 17, Appl
32	30	7.3	13613	4	US-09-105-537-3	Sequence 3, Appli
33	30	7.3	38506	3	US-09-320-878-19	Sequence 19, Appl
c 34	29.8	7.3	4177	3	US-09-023-082A-23	Sequence 23, Appl
35	29.8	7.3	4411529	4	US-09-103-840A-1	Sequence 1, Appli
36	29.6	7.2	35524	3	US-08-923-137-1	Sequence 1, Appli
37	29.4	7.2	815	4	US-08-818-112-139	Sequence 139, App
38	29.4	7.2	28958	1	US-08-258-261B-6	Sequence 6, Appli
39	29.4	7.2	28958	1	US-08-456-837-6	Sequence 6, Appli
40	29.4	7.2	28958	1	US-08-457-342-6	Sequence 6, Appli
41	29.4	7.2	28958	1	US-08-457-646A-6	Sequence 6, Appli
42	29.4	7.2	28958	1	US-08-458-076A-6	Sequence 6, Appli
43	29.4	7.2	28958	1	US-08-764-233A-4	Sequence 4, Appli
44	29.4	7.2	28958	1	US-08-457-335A-6	Sequence 6, Appli
45	29.4	7.2	28958	1	US-08-729-214-6	Sequence 6, Appli

ALIGNMENTS

```

RESULT      1
US-09-007-005-17/c
; Sequence 17, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; TITLE OF INVENTION: FUSIONS
; FILE REFERENCE: 00786/350003
; CURRENT APPLICATION NUMBER: US/09/007,005B
; CURRENT FILING DATE: 1998-01-14
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(289)

```

; OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17

Query Match 8.5%; Score 35; DB 4; Length 289;
Best Local Similarity 7.9%; Pred. No. 0.04;
Matches 20; Conservative 102; Mismatches 131; Indels 0; Gaps 0;

```
Qy      91 cccatcccaacactccaaaccaaccaacacttcaaccaaaccaccacaacaatgccttca 150
          |:|: : |:|: | |: |:| |:|: :: :: : : :: : :
Db      253 YCYAYAYGYAYGYTYTYAYCYGYCYAYGYCYTYGYSYNYNSYNYNSYNYNSYNYNS 194

Qy     151 gtaaccaggcccggtctcatgtggcgtagcgctcgcccgcgccatggttagagccccacccc 210
          ::: ::: : ::: : : : : : : : : : : : : : : : :
Db     193 YNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNS 134

Qy     211 ttcgctcgcaatcccatcaccatgacccctcacgcctggcgcgccgacctctccaag 270
          : : : : : : : : : : : : : : : : : : : : : : :
Db     133 YNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNS 74

Qy     271 aaagtcgtgaagacaagcactgtcttcttcccttctatgcaggtatccttggatggcca 330
          :: : | : : : |: : |: : |: : |: : |: : |: : :
Db       73 YNYNSYNYNSYNYNYCYAYTYTYGYTYAYAYTYTYGYTYAYAYAYTYAYGYTYAYAYT 14

Qy     331 gtcgcagccgcct 343
          |:|: |:|:
Db      13 YTYGYTYCYCYCY 1
```

RESULT 2
US-09-244-796-17/c
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; TITLE OF INVENTION: FUSIONS
; FILE REFERENCE: 00786/350007
; CURRENT APPLICATION NUMBER: US/09/244,796
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; EARLIER APPLICATION NUMBER: 09/007,005
; EARLIER FILING DATE: 1998-01-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; FEATURE:

; NAME/KEY: misc_feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17

Query Match 8.5%; Score 35; DB 4; Length 289;
Best Local Similarity 7.9%; Pred. No. 0.04;
Matches 20; Conservative 102; Mismatches 131; Indels 0; Gaps 0;

```
Qy      91 cccatcccaacactccaaaccaaccaacttcaaccaaaccaccacaacaatgccttca 150
          |:|: : |:|: | |: |:| |:|: :: :: : : ::: : :
Db      253 YCYAYAYGYAYGYTYTYAYCYGYCYAYGYCYTYGYSYNYNSYNYNSYNYNSYNYNS 194

Qy     151 gtaaccagggcccggtctcatgtggcgtagcgtagcgcccgcgcatggtagagccccacccc 210
          ::: ::: : ::: ::: : : ::: : : : : : : ::: : :
Db     193 YNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNS 134

Qy     211 ttcgctcgcaatcccatcaccatgacccctcacgcctggcgcgccgacacctctccaag 270
          : : : : : : : : : : : : : : : : : : : : : :
Db     133 YNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNSYNYNS 74

Qy     271 aaagtcgtgaagacaagcactgtcttcttcccttctatgcaggtatccttggatggcca 330
          :: : | : : : |: : |: : : |: : |: : : |: : :
Db       73 YNYNSYNYNSYNYNSYNYCYAYTYTYGYTYAYAYTYTYGYTYAYAYAYTYAYGYTYAYAYT 14

Qy     331 gtcgcagccgcct 343
          |:|: |:|:
Db      13 YTYGYTYCYCYCY 1
```

RESULT 3
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14

```

```

Query Match          8.5%; Score 35; DB 1; Length 7218;
Best Local Similarity 3.1%; Pred. No. 0.2;
Matches 11; Conservative 192; Mismatches 152; Indels 0; Gaps 0;

```

```

Qy    17 tccatcttctctgctcaatcaattacacaacaagagcattctagatttgagttcatccta 76
      ::: :::::::::: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db    1102 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1161

Qy    77 gcgataccaatacacccatcccaacactccaaaccaaccaacacttcaaccaaacacca 136
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db    1162 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1221

Qy    137 caacaatgccttcagtaaacccaggcccgctctcatgtggcgtagcgctcgccgcggcatgg 196
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db    1222 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1281

Qy    197 tagagccccacccttcgctcgcaatcccatcaccatgaccctcagcgctggcgcgccg 256
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db    1282 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1341

Qy    257 ccgacctctccaagaaagtcgtgaagacaagcactgtcttcttccccttctatgcaggta 316
      :: :::::::::: :: : : : : : : : : : : : : : : : : : : :
Db    1342 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1401

Qy    317 tccttggaaggccagtcgcagccgcctggtggttcaacggaaacatgtgactctt 371
      :::: : : : : : : : : : : : : : : : : : : : : : : : :
Db    1402 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYGTACCAAATTCTTCTATCTCTT 1456

```

```

RESULT 4
US-08-941-936-1
; Sequence 1, Application US/08941936
; Patent No. 6054305

```

```

; GENERAL INFORMATION:
;   APPLICANT:  Tatsumi, Hiroki
;   APPLICANT:  Eisaki, Naoki
;   APPLICANT:  Horiuchi, Tatsuo
;   APPLICANT:  Nagahara, Ayumu
;   TITLE OF INVENTION:  Pyruvate Orthophosphate Dikinase Gene,
;   TITLE OF INVENTION:  Recombinant DNA, and Process For Producing Pyruvate
;   TITLE OF INVENTION:  Orthophosphate Dikinase
;   NUMBER OF SEQUENCES:  8
;   CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  MEDLEN & CARROLL, LLP
;       STREET:    220 Montgomery Street, Suite 2200
;       CITY:      San Francisco
;       STATE:     CA
;       COUNTRY:   US
;       ZIP:       94104
;   COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:     IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:     PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/941,936
;       FILING DATE:       01-OCT-1997
;       CLASSIFICATION:    435
;   ATTORNEY/AGENT INFORMATION:
;       NAME:             Carroll, Peter G.
;       REGISTRATION NUMBER:  32,837
;       REFERENCE/DOCKET NUMBER:  HIRAKI-03009
;   TELECOMMUNICATION INFORMATION:
;       TELEPHONE:        415-705-8410
;       TELEFAX:          415-397-8338
;   INFORMATION FOR SEQ ID NO:  1:
;       SEQUENCE CHARACTERISTICS:
;           LENGTH:        2634 base pairs
;           TYPE:           nucleic acid
;           STRANDEDNESS:   double
;           TOPOLOGY:       linear
;       MOLECULE TYPE:      DNA (genomic)
;       ORIGINAL SOURCE:
;           ORGANISM:       Microbispora thermorosea
;           STRAIN:         IFO 14047
;       FEATURE:
;           NAME/KEY:       CDS
;           LOCATION:       1..2634
US-08-941-936-1

```

```

Query Match          8.3%;  Score 34.2;  DB 3;  Length 2634;
Best Local Similarity 55.5%;  Pred. No. 0.22;
Matches 66;  Conservative 0;  Mismatches 53;  Indels 0;  Gaps 0;

```

```

Qy  180  cgctgccccgcggcatggtagagccccacccttcgctcgcaatcccatcaccatgacccc 239
      ||| ||||| ||||| || || | || | || | || || ||
Db  1353  CGTGGCCCGCGGCATGGGCAAGACCTGCGTGTGCGGGGCCGAGGAAGTGAAGTGGACCC 1412

Qy  240  tcacgcctggcgcgccgacacctctccaagaaagtcgtgaagacaagcactgtcttct 298

```

||||| | ||| | ||| | | | || ||||| | | |||
Db 1413 GCACGCCCGCGCTTACCGCGCCCGGCGGGATCGTCGTGAACGAGGGCGAGGTGATCT 1471

RESULT 5
US-09-060-756-4
; Sequence 4, Application US/09060756
; Patent No. 6183957
; GENERAL INFORMATION:
; APPLICANT: Cole, Stewart
; APPLICANT: Buchrieser-Brosch, Roland
; APPLICANT: Gordon, Stephen
; APPLICANT: Billault, Alain
; TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
; TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
; TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA
; FILE REFERENCE: 3495-0169
; CURRENT APPLICATION NUMBER: US/09/060,756
; CURRENT FILING DATE: 1998-04-16
; NUMBER OF SEQ ID NOS: 743
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1280
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-09-060-756-4

Query Match 8.0%; Score 33; DB 4; Length 1280;
Best Local Similarity 47.8%; Pred. No. 0.38;
Matches 96; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

Qy 160 gcccgctctcatgtggcgtagcgctgccccgcgccatggttagagccccacccttcgctcgc 219
| | ||| | ||| ||||| || | |||| | | | | |||| |
Db 787 ggcgggtcagttatggggtagcggcgccgcgccgctcgaaggcggcgagccttaagcgtc 846

Qy 220 aatcccatcaccatgaccctcacgcctggcgcgccgacctctccaagaaagtctgtg 279
| | | | | | | | | | | | | | | | | | | | | |
Db 847 ggcgacaccggcgggggcggtggcgctcgccgagcgccgggctgatcgccaccggcgcc 906

Qy 280 aagacaagcactgtcttcttcccttctatgcaggtatccttgatggccagtcgcagcc 339
| | ||| | | | | | | | | | | | | | | | | | |
Db 907 aacggcggaacggcgccaccggcgccaacgcggcagccccggaaccggcgccggcgcc 966

Qy 340 gcctggtggttcaacggaaac 360
| | | | | | ||||
Db 967 gggttgctgctgggcaaaac 987

RESULT 6
US-08-858-003-1
; Sequence 1, Application US/08858003
; Patent No. 6060234
; GENERAL INFORMATION:
; APPLICANT: Katz, Leonard
; APPLICANT: Stassi, Diane L.
; APPLICANT: Summers Jr., Richard G.

```

; APPLICANT: Ruan, Xiaoan
; APPLICANT: Pereda-Lopez, Ana
; APPLICANT: Kakavas, Stephan J.
; TITLE OF INVENTION: NOVEL POLYKETIDE DERIVATIVES
; TITLE OF INVENTION: AND RECOMBINANT METHODS FOR MAKING SAME
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Rd.
; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/858,003
; FILING DATE: 16-MAY-1979
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dianne Casuto
; REGISTRATION NUMBER: P-40,943
; REFERENCE/DOCKET NUMBER: 4952.US.P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (847)-938-3137
; TELEFAX: (847)-938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 925 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-858-003-1

```

```

Query Match          8.0%; Score 32.8; DB 3; Length 925;
Best Local Similarity 49.4%; Pred. No. 0.37;
Matches 85; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

```

```

Qy   96 cccaacactccaaaccaaccaacacttcaaccaaaccaccacaacaatgccttcagtaac 155
      || || ||||| || ||| || |||| ||||| || || |||
Db   408 CCACACCCTCCAACCCACCTCGACAACCACCACGACACCATCTCCATCGCCGCCATCAA 467

Qy   156 ccaggcccggtctcatgtggcgtagcgtagcgccgcggcatggtagagccccacccttcgc 215
      | ||| || || | |||| || || || || || || |
Db   468 CGGCCCCCAGCCACCGTCTCTCCGGCGACCGCACCACCCTCCACCACATCGCCACCCA 527

Qy   216 tcgcaatcccatcaccatgaccctcacgcctggcgcgccgacgtctcc 267
      | ||| ||| || |||| | | | || ||| |||
Db   528 ACTCAACACCAAACCTTCACCACCACCCTCAACACCCTACCCACCACCCC 579

```

RESULT 7
 US-09-078-166-1
 ; Sequence 1, Application US/09078166
 ; Patent No. 6063561
 ; GENERAL INFORMATION:
 ; APPLICANT: Katz, Leonard
 ; APPLICANT: Stassi, Diane L.
 ; APPLICANT: Summers Jr., Richard G.
 ; APPLICANT: Ruan, Xiaoan
 ; APPLICANT: Pereda-Lopez, Ana
 ; APPLICANT: Kakavas, Stephan J.
 ; TITLE OF INVENTION: NOVEL POLYKETIDE DERIVATIVES
 ; TITLE OF INVENTION: AND RECOMBINANT METHODS FOR MAKING SAME
 ; NUMBER OF SEQUENCES: 44
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Abbott Laboratories
 ; STREET: 100 Abbott Park Rd.
 ; CITY: Abbott Park
 ; STATE: Illinois
 ; COUNTRY: USA
 ; ZIP: 60064-3500
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSEQ Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/078,166
 ; FILING DATE: 16-MAY-1979
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Dianne Casuto
 ; REGISTRATION NUMBER: P-40,943
 ; REFERENCE/DOCKET NUMBER: 4952.US.P2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (847)-938-3137
 ; TELEFAX: (847)-938-2623
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 925 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 US-09-078-166-1

Query Match 8.0%; Score 32.8; DB 3; Length 925;
 Best Local Similarity 49.4%; Pred. No. 0.37;
 Matches 85; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

Qy 96 cccaacactccaaaccaaccaacacttcaaccaaaccaccacaacaatgccttcagtaac 155


```

      || || ||||| || ||| || | |||| |||| | || | | |
Db    408 CCACACCCTCCAACCCACCTCGACAACCACCACGACACCATCTCCATCGCCGCCATCAA 467

Qy    156 ccaggcccggtctcatgtggcgtagcgtcgcccgcggcatggtagagccccacccttcgc 215
      | ||| || | || || |||| || | || | | || |
Db    468 CGGCCCCACGCCACCGTCCTCTCCGGCGACCGCACCACCCTCCACCACATCGCCACCCA 527

Qy    216 tcgcaatcccatcaccatgaccctcacgcctggcgcgccgacgtctcc 267
      | ||| ||| || || || | | | | || || || |
Db    528 ACTCAACACCAAACCTTCACCACCACCCTCAACACCCTCACCACCAACCCC 579

```

RESULT 8

US-08-997-467-1

; Sequence 1, Application US/08997467

; Patent No. 6200813

; GENERAL INFORMATION:

; APPLICANT: Katz, Leonard

; APPLICANT: Stassi, Diane L.

; APPLICANT: Summers Jr., Richard G.

; APPLICANT: Ruan, Xiaoan

; APPLICANT: Pereda-Lopez, Ana

; APPLICANT: Kakavas, Stephan J.

; TITLE OF INVENTION: NOVEL POLYKETIDE DERIVATIVES

; TITLE OF INVENTION: AND RECOMBINANT METHODS FOR MAKING SAME

; NUMBER OF SEQUENCES: 34

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Abbott Laboratories

; STREET: 100 Abbott Park Rd.

; CITY: Abbott Park

; STATE: Illinois

; COUNTRY: USA

; ZIP: 60064-3500

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSEQ Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/997,467

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/858,003

; FILING DATE: 16-MAY-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Dianne Casuto

; REGISTRATION NUMBER: P-40,943

; REFERENCE/DOCKET NUMBER: 4952.US.P2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (847)-938-3137

; TELEFAX: (847)-938-2623

; TELEX:

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 925 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-997-467-1

Query Match 8.0%; Score 32.8; DB 4; Length 925;
Best Local Similarity 49.4%; Pred. No. 0.37;
Matches 85; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

```
Qy 96 cccaacactccaaaccaacacacttcaaccaaaccaccacaacaatgccttcagtaac 155
    || || ||||| || ||| || | |||| ||||| || | | | |
Db 408 CCACACCCTCCAACCCACCTCGACAACCACCACGACACCATCTCCATCGCCGCCATCAA 467

Qy 156 ccaggcccgtctcatgtggcgtagcgctcgcccgcgcatggtagagccccaccccttcgc 215
    | ||| || | | || ||||| || | | | | | |
Db 468 CGGCCCCCAGGCCACCGTCTCTCCGGCGACCGCACCCCTCCACCACATCGCCACCCA 527

Qy 216 tcgcaatcccatcaccatgacccctcacgcctggcgcgccgacctctcc 267
    | ||| || | | || | | | | | | | | | | |
Db 528 ACTCAACACCAAACCCCTTACCACCACCCTCAACACCCTCACCCACCACCCC 579
```

RESULT 9

US-08-764-233A-1

; Sequence 1, Application US/08764233A

; Patent No. 5716849

; GENERAL INFORMATION:

; APPLICANT: Ligon, James M.

; APPLICANT: Schupp, Thomas

; APPLICANT: Beck, James J.

; APPLICANT: Hill, Dwight S.

; APPLICANT: Neff, Snezana

; APPLICANT: Ryals, John A.

; TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Ciba-Geigy Corporation

; STREET: 520 White Plains Road, P.O. Box 2005

; CITY: Tarrytown

; STATE: NY

; COUNTRY: USA

; ZIP: 10591

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/764,233A

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/729,214

; FILING DATE: 09-OCT-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/258,261

; FILING DATE: 08-JUN-1994

```

; ATTORNEY/AGENT INFORMATION:
;   NAME: Meigs, J. Timothy
;   REGISTRATION NUMBER: 38,241
;   REFERENCE/DOCKET NUMBER: 1506/CIP6
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (919) 541-8587
;   TELEFAX: (919) 541-8689
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 49377 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   ORIGINAL SOURCE:
;     ORGANISM: Sorangium cellulosum
;   IMMEDIATE SOURCE:
;     CLONE: p98/1, pJL3, and pVKM15
;   FEATURE:
;     NAME/KEY: misc_feature
;     LOCATION: 383..760
;     OTHER INFORMATION: /product= "SorR"
;     OTHER INFORMATION: /note= "This gene encodes a protein that is highly
homologous t
;     OTHER INFORMATION: the reductase domains of type I PKSs such as eryA
from
;     OTHER INFORMATION: Saccharopolyspora erythraea."
;   FEATURE:
;     NAME/KEY: misc_feature
;     LOCATION: 927..19874
;     OTHER INFORMATION: /product= "SorA"
;     OTHER INFORMATION: /note= "Gene product is highly homologous to type I
PKSs that
;     OTHER INFORMATION: are known to be involved in the synthesis of
polyketide
;     OTHER INFORMATION: compounds."
;   FEATURE:
;     NAME/KEY: misc_feature
;     LOCATION: 942..7115
;     OTHER INFORMATION: /product= "Module 1 of SorA"
;   FEATURE:
;     NAME/KEY: misc_feature
;     LOCATION: 7203..12884
;     OTHER INFORMATION: /product= "Module 2 of SorA"
;   FEATURE:
;     NAME/KEY: misc_feature
;     LOCATION: 13455..19616
;     OTHER INFORMATION: /product= "Module 3 of SorA"
;   FEATURE:
;     NAME/KEY: misc_feature
;     LOCATION: 19871..46318
;     OTHER INFORMATION: /product= "SorB"
;     OTHER INFORMATION: /note= "Gene product is highly homologous to type I
PKS genes."
;   FEATURE:
;     NAME/KEY: misc_feature
;     LOCATION: 19870..24556

```

```

;   OTHER INFORMATION:  /product= "Module 1 of SorB"
;   FEATURE:
;   NAME/KEY:  misc_feature
;   LOCATION:  24638..30820
;   OTHER INFORMATION:  /product= "Module 2 of SorB"
;   FEATURE:
;   NAME/KEY:  misc_feature
;   LOCATION:  30881..35446
;   OTHER INFORMATION:  /product= "Module 3 of SorB"
;   FEATURE:
;   NAME/KEY:  misc_feature
;   LOCATION:  35528..40114
;   OTHER INFORMATION:  /product= "Module 4 of SorB"
;   FEATURE:
;   NAME/KEY:  misc_feature
;   LOCATION:  40190..46318
;   OTHER INFORMATION:  /product= "Module 5 of SorB"
;   FEATURE:
;   NAME/KEY:  misc_feature
;   LOCATION:  46851..47891
;   OTHER INFORMATION:  /product= "SorM"
;   OTHER INFORMATION:  /note= "The protein encoded by the sorM gene is
highly
;   OTHER INFORMATION:  homologous to the methyltransferase from Streptomyces
;   OTHER INFORMATION:  hygrosopicus that is involved in the synthesis of
the
;   OTHER INFORMATION:  polyketide rappamycin."
US-08-764-233A-1

```

```

Query Match          7.9%;  Score 32.6;  DB 1;  Length 49377;
Best Local Similarity 46.3%;  Pred. No. 3.2;
Matches 107;  Conservative 0;  Mismatches 124;  Indels 0;  Gaps 0;

```

```

Qy   101 cactccaaaccaaccaacacttcaaccaaaccaccacaacaatgccttcagtaaccagg 160
      | ||||| || | || || || | | | | | | | | | | | | | |
Db 12535 CCCTCCAAGCCCTCTTGGACTCCATCCCAGCGCTCACCCGCTACCGCCGTCGTCCACG 12594

Qy   161 cccgtctcatgtggcgtagcgctcgcccgggcatggtagagccccaccccttcgctcgca 220
      || | | | | | | | | | | | | | | | | | | | |
Db 12595 CCGCGGGCGCCCTCGACGACGGCCTGCTCGGCGCCATGAGCCCCGAGCGCATCGACCGCG 12654

Qy   221 atcccatcaccatgacccctcaagcctggcgcgccgacacctctccaagaaagtcgtga 280
      | || | | | | | | | | | | | | | | | | | | |
Db 12655 TCTTTGCCCCAAGCTCGATGCTGCTTGGCACTTGTCATGAGCTACCCAAGACAAGCCCC 12714

Qy   281 agacaagcactgtcttcttcccttctatgcaggtatccttggtggccag 331
      | | | | | | | | | | | | | | | | | | | |
Db 12715 TCGCCGCCTTCGTCTCTTCTCGTCCGCTGCTGGCGTCCTTGGTAGTCCAG 12765

```

```

RESULT 10
US-08-998-416-882/c
; Sequence 882, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter

```

```

; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Rebischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 882:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 490 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1552UP
US-08-998-416-882

```

```

Query Match          7.6%; Score 31.4; DB 4; Length 490;
Best Local Similarity 48.5%; Pred. No. 0.77;
Matches 83; Conservative 1; Mismatches 87; Indels 0; Gaps 0;

```

```

Qy 209 ccttogctogcaatcccatcaccatgacccctcacgcctggcgcgccgcccacctctcca 268
    || || ||| | || |||| ||| |||| |||| ||| | ||
Db 356 CCGTCATCCGCGACGCCGTACCTACACCGAGCACGCCAAGCGCAAGACCGTCACCTCGC 297

Qy 269 agaaagtcgtgaagacaagcactgtcttcttccccttctatgcaggtatccttggatggc 328
    | ||||| | | || | || | | || || || ||

```



```

; REGISTRATION NUMBER: 31,145
; REFERENCE/DOCKET NUMBER: 31,255-02
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201)831-3241
; TELEFAX: (201)831-3305
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30001 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-474-933-1

```

Query Match 7.6%; Score 31.2; DB 2; Length 30001;
Best Local Similarity 45.0%; Pred. No. 7.1;
Matches 117; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

Qy	78	cgataccaatacaccccatcccaaacactccaaaccaaccaacacttcaaccaaaccaccac	137
Db	8059	CGTGATCGACACCAACCTCACCAGCGTCTTCCGCGTCACCCGCGAGGTCCTCACCACCGG	8000
Qy	138	aacaatgccttcagtaacccaggcccgctctcatgtggcgtagcgctcgcccgggcatggt	197
Db	7999	CGGCATGGAGGCGGCCGGGCACGGGCGGATCATCAGCGTCGCTCCACCGGCGGCAAGCA	7940
Qy	198	agagccccaccccttcgctcgcaatcccatcaccatgacccctcacgcctggcgcgccgc	257
Db	7939	GGGTGTCCCGCTGGGCGCCCCCTACTCGGCCCTCCAAGGCCGGCGTCATCGGCTTCACCA	7880
Qy	258	cgacctctccaagaaagtcgtgaagacaagcactgtcttcttccccttctatgcaggtat	317
Db	7879	GGCGCTGGCCAAGGAACCTCGCCACACCGGCACCACGGTCAACGCCGTCTGCCCCGGCTA	7820
Qy	318	ccttggatggccagtcgcag	337
Db	7819	CGTCGAGACGCCGATGGCCG	7800

```

RESULT 13
US-09-056-105-9
; Sequence 9, Application US/09056105
; Patent No. 6287569
; GENERAL INFORMATION:
; APPLICANT: KIPPS, THOMAS J.
; APPLICANT: WU, YUNQI
; TITLE OF INVENTION: VACCINES WITH ENHANCED INTRACELLULAR
; TITLE OF INVENTION: PROCESSING
; FILE REFERENCE: 233/221
; CURRENT APPLICATION NUMBER: US/09/056,105
; CURRENT FILING DATE: 1998-04-06
; EARLIER APPLICATION NUMBER: 60/043,467
; EARLIER FILING DATE: 1997-04-10
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 9

```

; Sequence 9; Application US/09056105

; Patent No. 6287569

; GENERAL INFORMATION:

; APPLICANT: KIPPS, THOMAS J.

; APPLICANT: WU, YUNQI

; TITLE OF INVENTION: VACCINES WITH ENHANCED INTRACELLULAR

; TITLE OF INVENTION: PROCESSING

; FILE REFERENCE: 233/221

; CURRENT APPLICATION NUMBER: US/09/056,105

; CURRENT FILING DATE: 1998-04-06

; EARLIER APPLICATION NUMBER: 60/043,467

; EARLIER FILING DATE: 1997-04-10

; NUMBER OF SEQ ID NOS: 35

```
; SOFTWARE: FastSEQ for Windows Version 3.0
```

; SEQ ID NO 9

; LENGTH: 11495
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-056-105-9

Query Match 7.5%; Score 30.8; DB 4; Length 11495;
Best Local Similarity 49.4%; Pred. No. 5.9;
Matches 80; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

```
Qy      72  tcctagcgataccaatacacccatcccaacactccaaaccaaccaacacttcaaccaaac 131
      |||  |||  ||  ||  |||  ||  |||  ||  |||  |||  |||  |||  |||  |||
Db      604  tcccttccacactcctaaccaatccacaccctcatcccctaccagcaccccatcctccc 663

Qy     132  caccacaacaatgccttcagtaacccaggcccgtctcatgtggcgtagcgtcgcccgcg 191
      ||  ||  ||  |||  ||  |||  |||  |||  |||  ||  ||  ||  ||  ||  ||
Db     664  caaccccgtagccaccctcatacccccatccccaattcaaccccgcacccctcatcccca 723

Qy     192  catggtagagccccaccccttcgctcgcaatcccataccat 233
      |  |  |  |||  |  |  ||  |  |  |||  |||
Db     724  cccacacctgcacccccaccccccaacaccatacccccat 765
```

RESULT 14

US-08-752-760A-1

; Sequence 1, Application US/08752760A

; Patent No. 5877011

; GENERAL INFORMATION:

; APPLICANT: Armentano, Donna

; APPLICANT: Gregory, Richard J.

; APPLICANT: Smith, Alan E.

; TITLE OF INVENTION: CHIMERIC ADENOVIRAL VECTORS

; NUMBER OF SEQUENCES: 3

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Baker & Botts, L.L.P.

; STREET: 30 Rockefeller Plaza

; CITY: New York

; STATE: NY

; COUNTRY: U.S.A.

; ZIP: 10112

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSEQ Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/752,760A

; FILING DATE: 20-NOV-1996

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Seide, Rochelle K

; REGISTRATION NUMBER: 32,300

; REFERENCE/DOCKET NUMBER: A31385

; TELECOMMUNICATION INFORMATION:

```

; TELEPHONE: 212-705-5000
;
; TELEFAX: 212-705-5020
;
; TELEX:
;
; INFORMATION FOR SEQ ID NO: 1:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 35081 base pairs
;
; TYPE: nucleic acid
;
; STRANDEDNESS: single
;
; TOPOLOGY: linear
;
US-08-752-760A-1

```

Query Match 7.4%; Score 30.6; DB 2; Length 35081;
Best Local Similarity 48.6%; Pred. No. 12;
Matches 84; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

Qy	96	cccaacactccaaaccaacacacttcaaccaaacaccacaacaatgccttcagtaac	155
Db	16373	CCGACCCCTGGCTCCAGCCTCCACCGCTACCCTTCCACTTCTACCGTCGCCACGGTCAC	16432
Qy	156	ccaggcccgctctcatgtggcgtagcgtagcccgcgggcatggtagagccccaccccttcgc	215
Db	16433	CGAGCCTCCCAGGAGGCGAAGATGGGGCCCCGCCAACCGGCTGATGCCCAACTACGTGTT	16492
Qy	216	tgcgaatcccatcaccatgacccctcaagcctggcgcgccgacgacgtctcca	268
Db	16493	GCATCCTTCCATTATCCCAGCGCCGGGCTACCGCGGCACCCGGTACTACGCCA	16545

```

RESULT 15
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

```

Query Match 7.4%; Score 30.6; DB 4; Length 4403765;

Best Local Similarity 57.6%; Pred. No. 48;
Matches 57; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

Qy 183 cgcccgcgcatggttagagccccaccccttcgctcgcaatcccatcaccatgacccctca 242
| | | | | | | | | | | | | | | | | | | | | |
Db 2765925 caccaccgtggtgttcgacgcactccccggcgccgacacggtcatcgacatctccgcca 2765984

Qy 243 cgcttggcgcgcgcgcgcacctctccaagaaagtcgtgaa 281
| | | | | | | | | | | | | | | | | | | | | |
Db 2765985 caccgtgcgcgcgcgcgcacgcctcaacgaccaagacctgga 2766023

Search completed: February 7, 2002, 11:22:20
Job time: 7906 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 08:20:54 ; Search time 4942.22 Seconds
(without alignments)
893.630 Million cell updates/sec

Title: US-09-394-745-6603
Perfect score: 411
Sequence: 1 agcaaaagcatagagatcca.....aggagaagaggaagggaccg 411

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*

```

13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	%		Query		DB		ID	Description
No.	Score	Match	Length	DB	ID			
	1	53.4	13.0	435	11	BI190440	BI190440	ile09fs.r
c	2	52.6	12.8	884	13	CNS006U0	AL065923	Drosophil
	3	52	12.7	939	13	CNS00CNG	AL059400	Drosophil
c	4	51.2	12.5	895	13	CNS0071A	AL066286	Drosophil
	5	46.8	11.4	996	11	BG333443	BG333443	602430365
c	6	44.2	10.8	1101	13	CNS00LT2	AL078714	Drosophil
	7	43.6	10.6	494	10	AA415063	AA415063	Mg0008 RC
	8	43.6	10.6	830	11	BF207170	BF207170	601870887
	9	43.6	10.6	1013	13	CNS016KT	AL106871	Drosophil
c	10	42.8	10.4	922	13	CNS0073W	AL066784	Drosophil
	11	42	10.2	1101	13	CNS016LW	AL106910	Drosophil
	12	41.8	10.2	748	13	CNS011EQ	AL100172	Drosophil
	13	41.6	10.1	1028	13	CNS01FKL	AL141958	Anopheles
	14	41.6	10.1	1344	11	BF973142	BF973142	602242133
	15	41.4	10.1	1737	11	BF868167	BF868167	963101A02
c	16	41	10.0	812	13	AZ183515	AZ183515	SP_1001_A
	17	41	10.0	1139	11	BF859156	BF859156	963001C08
	18	40.8	9.9	632	11	BG320266	BG320266	Zm03_03g0
c	19	40.6	9.9	852	10	AL563901	AL563901	AL563901
	20	40.6	9.9	1101	13	CNS00LO0	AL068607	Drosophil
c	21	40.4	9.8	707	13	AZ199083	AZ199083	SP_1038_B
c	22	40.4	9.8	1064	13	CNS016P9	AL107031	Drosophil
c	23	40.4	9.8	1101	13	CNS00FXE	AL071370	Drosophil
	24	40.2	9.8	692	13	B21282	B21282	T20G4-T7.1
	25	40.2	9.8	860	13	CNS018FL	AL109275	Drosophil
c	26	40.2	9.8	916	13	CNS015SQ	AL105860	Drosophil
	27	40.2	9.8	1003	11	BG180518	BG180518	602331443
c	28	40	9.7	767	13	AZ196564	AZ196564	SP_1032_A
	29	40	9.7	932	13	CNS06KYG	AL403502	T7 end of
c	30	39.8	9.7	772	11	BF627771	BF627771	HVSMEb000
	31	39.8	9.7	902	13	CNS03MHM	AL250771	Tetraodon
	32	39.8	9.7	980	10	AL536103	AL536103	AL536103
c	33	39.8	9.7	1101	13	CNS00ZZK	AL098330	Drosophil
c	34	39.6	9.6	528	11	BG786286	BG786286	SEAUMC006
	35	39.6	9.6	952	13	CNS0766T	AL431019	T7 end of
	36	39.6	9.6	1101	13	CNS016XR	AL107337	Drosophil
	37	39.4	9.6	638	13	CNS046XY	AL277279	Tetraodon
	38	39.4	9.6	1046	13	CNS07BWV	AL438437	T7 end of

c	39	39.4	9.6	1063	13	CNS040FA	AL268831	Tetraodon
c	40	39.4	9.6	1101	13	CNS017ZR	AL108705	Drosophil
c	41	39.2	9.5	363	10	AL534163	AL534163	AL534163
	42	39	9.5	544	13	CNS015XA	AL106024	Drosophil
c	43	39	9.5	798	13	CNS02PA9	AL207738	Tetraodon
c	44	39	9.5	823	10	AL573901	AL573901	AL573901
	45	39	9.5	880	11	BF298281	BF298281	016PbB02

ALIGNMENTS

RESULT 1
BI190440
LOCUS BI190440 435 bp mRNA EST 10-JUL-2001
DEFINITION ile09fs.r1 Fusarium sporotrichioides Tri 10 overexpressed cDNA library Fusarium sporotrichioides cDNA clone ile09fs 5', mRNA sequence.
ACCESSION BI190440
VERSION BI190440.1 GI:14664119
KEYWORDS EST.
SOURCE Fusarium sporotrichioides.
ORGANISM Fusarium sporotrichioides
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreales; mitosporic Hypocreales; Fusarium.
REFERENCE 1 (bases 1 to 435)
AUTHORS Ren,Q., Tag,A., Peplow,A., Lai,H., Kupfer,C., Peterson,A., Beremand ,M. and Roe,B.
TITLE Analysis of a Fusarium sporotrichioides EST database
JOURNAL Unpublished (2001)
COMMENT Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
Department of Chemistry and Biochemistry
Advanced Center for Genome Technology, University of Oklahoma
620 Parrington Oval, Norman, OK 73019, USA
Tel: 405 325 4912
Fax: 405 325 7762
Email: broe@ou.edu
Contact Dr. Marian Beremand regarding clone availability Included is the best homolog from a blastx search of Genbank nr 04-09-01 155 6e-10 gi|12718428|emb|CAC2 (AL513462) putative protein [Neurosporacra
Seq primer: T3
High quality sequence stop: 395.
FEATURES Location/Qualifiers
source 1. .435
/organism="Fusarium sporotrichioides"
/strain="Tri 10"
/db_xref="taxon:5514"
/clone="ile09fs"
/clone_lib="Fusarium sporotrichioides Tri 10 overexpressed cDNA library"
/note="Vector: pBlueScript SK-; Site_1: EcoRI; Site_2: XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript ; 3' end of cDNA cloned into XhoI site of pBluescript"
BASE COUNT 107 a 138 c 76 g 114 t
ORIGIN

Query Match 13.0%; Score 53.4; DB 11; Length 435;
 Best Local Similarity 55.7%; Pred. No. 0.0018;
 Matches 102; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

```

Qy 184 gcccgcggcatggtagagccccacccttcgctcgcaatcccatcaccatgaccctcac 243
    ||| || | | | | | | | | | | | | | | | | | | | | | |
Db 189 GCCGGCCGAGCTATGGAGTCTCACCCCTTCGAGCGTATTCCCCTCACTCAGAAGCCTGCT 248

Qy 244 gcctggcgcgccgcgcgacctctccaagaaagtcgtgaagacaagcactgtcttcttcccc 303
    | | | | | | | | | | | | | | | | | | | | | |
Db 249 TCTCCTGATTACGCCAAGATGTTCAAGCGAGTTGGCAGCCAGGCCCTCTTCTTCTCCCT 308

Qy 304 ttctatgcaggtatccttggatggccagtcgcagccgcctggtggttcaacggaaacatg 363
    || || | | | | | | | | | | | | | | | | | |
Db 309 GGCTTCGCTGTCATCCTTGGCTGGCCTTTGGCTGCCAGTATGCCTTTGACGGTAAACTG 368

Qy 364 tga 366
    | |
Db 369 TAA 371
  
```

RESULT 2
 CNS006U0/c
 LOCUS CNS006U0 884 bp DNA GSS 03-JUN-1999
 DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC #
 BACR14N21 of RPCI-98 library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.
 ACCESSION AL065923
 VERSION AL065923.1 GI:4944891
 KEYWORDS GSS.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 884)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 COMMENT Determination of this BAC-end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila
 melanogaster genome using these BACs. For further information
 please see <http://www.fruitfly.org> The BDGP Drosophila
 melanogaster BAC library was prepared by Kazutoyo Osoegawa and
 Aaron Mammoser in Pieter de Jong's laboratory in the Department of
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
 NY. The library is named RPCI-98 and was constructed by partial
 EcoRI digestion of Drosophila DNA provided by the BDGP from the
 isogenic strain y2; cn bw sp, the same strain used for the BDGP's
 P1 and EST libraries. A more detailed description of the library
 and how to order individual BAC clones, the entire library, or
 filters for hybridization from the BACPAC Resource Center can be
 found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

```

FEATURES                      Location/Qualifiers
  source                      1. .884
                              /organism="Drosophila melanogaster"
                              /db_xref="taxon:7227"
                              /clone_lib="RPCI-98"
                              /clone="BACR14N21"
                              /note="end : T7"
BASE COUNT      230 a      62 c      139 g      124 t      329 others
ORIGIN

```

```

Query Match      12.8%;  Score 52.6;  DB 13;  Length 884;
Best Local Similarity  15.6%;  Pred. No. 0.0034;
Matches  40;  Conservative 119;  Mismatches  97;  Indels    0;  Gaps    0;

```

```

Qy      18 ccatcttctctgctcaatcaattacacaacaagagcattctagatttgagttcatcctag 77
      :::| :: : | : :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db      876 MMHTKKKKKKTTHMMVMNMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMGMMMMMMMMMMMMMAMMMMMMM 817

Qy      78 cgataccaatacacccatcccaacactccaaaccaaccaacacttcaaccaaaccaccac 137
      : : :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db      816 MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMAMMCMMMCMMACCMCMAMMCAMMMMMCMMM 757

Qy     138 aacaatgccttcagtaaccaggcccgctctcatgtggcgtagcgctcgcccgcgccatggt 197
      :: :: : : || :: | : | : : : | : | : : : : :
Db      756 MMMMMCMMMMMCMCMCMCCMMCMCCMMMACACMACACMCACCMMMMACSMMAMC 697

Qy     198 agagccccacccttcgctcgcaatcccatcaccatgaccctcacgcctggcgcgccgc 257
      : : : | :: :: : : || :: || | :: || | | | : | : | | | |
Db      696 MCMCMACMMMMMMMMMAAMMMCCACHMCCAMACCMMMCCCCCMCACCMCMMMCCCCCCC 637

Qy     258 cgacctctccaagaaa 273
      | :: : | || : || :
Db      636 CMMMMACAACAMCAAM 621

```

```

RESULT      3
CNS00CNG
LOCUS      CNS00CNG      939 bp      DNA      GSS      04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
              BACR26H16 of RPCI-98 library from Drosophila melanogaster (fruit
              fly), genomic survey sequence.
ACCESSION  AL059400
VERSION    AL059400.1  GI:4946964
KEYWORDS   GSS.
SOURCE     fruit fly.
  ORGANISM Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE  1 (bases 1 to 939)
  AUTHORS  Genoscope.
  TITLE    Direct Submission
  JOURNAL  Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT    Determination of this BAC-end sequence was carried out as part of a

```

collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

```
FEATURES             Location/Qualifiers
    source             1. .939
                       /organism="Drosophila melanogaster"
                       /db_xref="taxon:7227"
                       /clone_lib="RPCI-98"
                       /clone="BACR26H16"
                       /note="end : TET3"
BASE COUNT           71 a    349 c    104 g    180 t    235 others
ORIGIN
```

```
Query Match          12.7%;  Score 52;  DB 13;  Length 939;
Best Local Similarity 16.0%;  Pred. No. 0.0049;
Matches 40;  Conservative 119;  Mismatches 91;  Indels 0;  Gaps 0;
```

```
Qy      18 ccattcttctctgctcaatcaattacacaacaagagcattctagatttgagttcatcctag 77
      :: ::| | : : :: :: : :: :: :: :: ::| :: : | | : : |::
Db      378 MMCHMTCCMMMMMMMMMMMMMMMMMTMHMMMMMMMMMMMMTMMHMMMMTTMMHAMTMMCMTMT 437

Qy      78 cgataccaatacacccatcccaacactccaaaccaacacacttcaaccaaaccaccac 137
      : : :::: |:: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Db      438 MMMMMMMMTTMMTMMMMMMMMMMMMMMMMTMMMMMMMMTMMMMMMMMTMMHMMMMMMHHMT 497

Qy      138 aacaatgccttcagtaacccaggcccgctctcatgtggcgtagcgctcgccgcggcgatggt 197
      :::: :: : :::: :: : |: :: : |: :: :: : :
Db      498 MMMMMHMMMMMMHTHHMMMMMMMMMTMMMTMMMMMMMMMMMMTMMMHMMMMMMMMMMHHMM 557

Qy      198 agagccccacccttcgctcgcaatcccatcaccatgaccctcacgcctggcgcgccgc 257
      : : |::: :: | | | | | | | | | | | | | | | |
Db      558 MMMMCMMMMMTTMTCCGTCTCTCTCTCCCGTGCCCTGCCCTTCTTGGTTGGCCCGTCTG 617

Qy      258 cgacctctcc 267
      | | | | |
Db      618 GGCCCTCCCC 627
```

```
RESULT      4
CNS0071A/c
LOCUS      CNS0071A      895 bp      DNA      GSS      03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR14B09 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION  AL066286
```



```

VERSION      AL066286.1   GI:4945153
KEYWORDS     GSS.
SOURCE       fruit fly.
ORGANISM      Drosophila melanogaster
              Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
              Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
              Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE    1 (bases 1 to 895)
AUTHORS      Genoscope.
TITLE        Direct Submission
JOURNAL      Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
              BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
              - Web : www.genoscope.cns.fr)
COMMENT      Determination of this BAC-end sequence was carried out as part of a
              collaboration with the Berkeley Drosophila Genome Project (BDGP).
              The BDGP is constructing a physical map of the Drosophila
              melanogaster genome using these BACs. For further information
              please see http://www.fruitfly.org The BDGP Drosophila
              melanogaster BAC library was prepared by Kazutoyo Osoegawa and
              Aaron Mammoser in Pieter de Jong's laboratory in the Department of
              Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
              NY. The library is named RPCI-98 and was constructed by partial
              EcoRI digestion of Drosophila DNA provided by the BDGP from the
              isogenic strain y2; cn bw sp, the same strain used for the BDGP's
              P1 and EST libraries. A more detailed description of the library
              and how to order individual BAC clones, the entire library, or
              filters for hybridization from the BACPAC Resource Center can be
              found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.
FEATURES      Location/Qualifiers
              source          1. .895
                               /organism="Drosophila melanogaster"
                               /db_xref="taxon:7227"
                               /clone_lib="RPCI-98"
                               /clone="BACR14B09"
                               /note="end : TET3"
BASE COUNT    124 a          80 c          204 g          179 t          308 others
ORIGIN

```

Qy	4	aaaagcatagagatccatcttctctgctcaatcaattacacaacaagagcattttagatt	63
		::: : : : : :: : ::: : : :	
Db	650	MMAMCCAMMCACAMMAMCMMCCCCACMMCACCCMMMCMCAMAMMAMCACMMCACMACM	591
Qy	64	tgagttcatcctagcgataccaatacacccatcccaacactccaaaccaaccaacacttc	123
		: : :: :: ::: : : ::: : :: : :	
Db	590	MMAMCMCAMMMAMCMAMMMMAAMACACMMMCCAMAMMCMMMMMMACMCACMMCCAMMAM	531
Qy	124	aaccaaaaccaccacaacaatgccttcagtaaccaggcccgtctcatgtggcgtagcgtc	183
		::: ::: :: ::: : :::: : :: :	
Db	530	MMMMAMMMCCMMMCMCMCMCCCCMMMMMMMMMAMCCAMMAAMMCMCMCMCCCMCMAC	471
Qy	184	gcccgcggcatggtagagccccacccttcgctcgcaatcccatcaccatgaccctcac	243
		: :: : : : :	

Matches 105; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

```
Qy      66 agttcatcctagcgataccaatacacccatcccaacactccaaaccaaccaacacttcaa 125
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      783 ATTGCCCCCTCCCCCTCCCAACCAACCCACCTCCCCACATCTGTCCCAACCCCACTCAC 842

Qy     126 ccaaaccaccacaacaatgccttcagtaaccaggcccggtctcatgtggcgtagcgtcgc 185
      || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     843 CCCCCTACACCACCTAACCCCTCATCACCCCAACCCCAACCCCAACGCCCCCTCCCCCAC 902

Qy     186 ccgcggcgatggtagagccccaccccttcgctcgcaatcccatcaccatgacccctcacgc 245
      || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     903 CCCCCACCTCCCCCACACCCACCCCTCCCCCAACCCCAACCCCAACCTTCCCTCCCCCCCC 962

Qy     246 ctggcgcgccgacgtctcc 267
      | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     963 CCCCTCCCCCCCCCCCCCCCCCCC 984
```

RESULT 6

CNS00LT2/c

LOCUS CNS00LT2 1101 bp DNA GSS 14-JUN-1999

DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC:
BACR48P19 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.

ACCESSION AL078714

VERSION AL078714.1 GI:5102004

KEYWORDS GSS.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1101)

AUTHORS Genoscope.

TITLE Direct Submission

JOURNAL Submitted (11-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)

COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES Location/Qualifiers

source 1. .1101

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR48P19"
/note="end : TET3"

BASE COUNT 469 a 6 c 69 g 151 t 406 others
ORIGIN

Query Match 10.8%; Score 44.2; DB 13; Length 1101;
Best Local Similarity 13.0%; Pred. No. 0.48;
Matches 39; Conservative 141; Mismatches 121; Indels 0; Gaps 0;

```
Qy      3  caaaagcatagagatccatcttctctgctcaatcaattacacaacaagagcattctagat 62
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db    1032  MMMMMMMMAMMMAMMMAMMMAMMMATTTTTHMMAMAMMMHMMMMMMMMMATTHAWHTTTTHT 973

Qy     63  ttgagttcatcctagcgataccaatacacccatcccaacactccaaaccaaccaacactt 122
      | | :    | : | : | : : : : | : : | : : | : : | : : | : : | : : |
Db     972  TTYMMAMCMTTHTMMMMMMMMMAMMAMMMCCMCMCMCCMMMMCCCCMCMMMMMCMMMTT 913

Qy    123  caaccaaaccaccacaacaatgccttcagtaacccaggcccggtctcatgtggcgtagcgt 182
      : : : : | : : | : : : : : : : : : : : : : : : : : : : : :
Db     912  TTHHHMMCCMCMCMCCCMCMCMCMMMMACMMMMTTTTHMMCCMMMMMMMMMMMAAMMMA 853

Qy    183  cgcccgcgcatggtagagccccaccccttcgctcgcaatcccatcaccatgacccctca 242
      : : : : : : | : : : : : : : : : : : : : : : : : : : : :
Db     852  MMMMMCMHMAMMTTTTTTTTTHWMAMMAYHTTMMMTTMMMMMMCMMMMAAAATTMMMMMM 793

Qy    243  cgccctggcgcgccgacacctctccaagaaagtcgtgaagacaagcactgtcttcttccc 302
      : : | : : : : : | | : : : : : : : : : : : : : : : : : :
Db     792  AMHMTTHHCTMMCMCMCWCCMCMMMMMMMCCMMCMCMCMCCMMYHMMHTTTWTMM 733

Qy    303  c 303
      :
Db     732  H 732
```

RESULT 7
AA415063
LOCUS AA415063 494 bp mRNA EST 09-DEC-1999
DEFINITION Mg0008 RCW Lambda Zap Express Library Magnaporthe grisea cDNA clone
 RCW8 3', mRNA sequence.
ACCESSION AA415063
VERSION AA415063.1 GI:2537242
KEYWORDS EST.
SOURCE Magnaporthe grisea.
 ORGANISM Magnaporthe grisea
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
REFERENCE 1 (bases 1 to 494)
 AUTHORS Wu, S.-C., Bernstein, B.D., Darvill, A.G. and Albersheim, P.
 TITLE Expressed sequence tags of the rice blast fungus grown on rice cell
 walls
 JOURNAL Unpublished (1997)
COMMENT Contact: Sheng-Cheng Wu
 CCRC
 University of Georgia

220 Riverbend Road, Athens, GA 30602-4712, USA

Tel: 706 542 4446

Fax: 706 542 4412

Email: wusc@bscr.uga.edu

Identical to Mg0040, Mg0046

Seq primer: T7.

FEATURES
source Location/Qualifiers
1. .494
/organism="Magnaporthe grisea"
/strain="CP987"
/db_xref="taxon:148305"
/clone="RCW8"
/clone_lib="RCW Lambda Zap Express Library"
/tissue_type="Mycelium"
/dev_stage="Day 5 post-inoculation"
/note="Vector: Lambda Zap; Messenger RNAs prepared from
Magnaporthe grisea grown at 23C in the dark with constant
gyratory shaking (100 rpm) in Vogel's medium containing
0.5% isolated rice cell walls as the sole carbon source"
BASE COUNT 136 a 124 c 115 g 119 t
ORIGIN

Query Match 10.6%; Score 43.6; DB 10; Length 494;
Best Local Similarity 52.8%; Pred. No. 0.56;
Matches 94; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

Qy 197 tagagccccaccccttcgctcgcaatcccatcaccatgacccctcacgcctggcgcgccg 256
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 35 TTGAGCCCCATCCCTTCCAGCGTCTGCCAACACGCAGAACGCCGCAAGCTGCCGACTATG 94
Qy 257 ccgacctctccaagaaagtcgtgaagacaagcactgtcttcttcccttctatgcaggta 316
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 95 CCAAGATCTTCAGGCGTTCGGCAAGACTGTCATGATCTACTTCCCTGGCATGGCCTTGA 154
Qy 317 tccttggtatggccagtcgcagccgcctggtggttcaacggaaacatgtgactcttcca 374
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 155 TCCTAGGTTGGCCTGTGATCGCTCAGAAGATGGTTGATGGCCACGTCTAAGGTCGCCA 212

RESULT 8
BF207170
LOCUS BF207170 830 bp mRNA EST 06-NOV-2000
DEFINITION 601870887F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4101692 5',
mRNA sequence.
ACCESSION BF207170
VERSION BF207170.1 GI:11100756
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 830)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM973 row: m column: 21
High quality sequence stop: 343.

FEATURES Location/Qualifiers
source 1. .830
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4101692"
 /clone_lib="NIH_MGC_19"
 /tissue_type="neuroblastoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

BASE COUNT 145 a 460 c 152 g 73 t
ORIGIN

Query Match 10.6%; Score 43.6; DB 11; Length 830;
Best Local Similarity 52.2%; Pred. No. 0.64;
Matches 97; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

```

Qy      82  accaatacacccatcccaacactccaaaccaacacacttcaaccaaaccaccacaaca 141
      ||||  | |||  ||| || |  || |||  ||||  | |  |  ||  ||||  ||
Db      590  ACCACCCCCCCCCCCCCACCCACACACCCACCACCCCAACCCCGCCCCATCACACCA 649

Qy      142  atgccttcagtaacccaggcccggtctcatgtggcgtagcgctcgcccgcgccatggttagag 201
      |||  |  ||||  ||  ||  |  ||  |  |  |  |  |  |  |
Db      650  TCACCTCAGCCACCCACCAGCCCGCCCCCGCCCCCGCCACCACCCCCCCCCACCA 709

Qy      202  cccacccccttcgctcgcaatcccatcaccatgacccctcacgcctggcgcgccgccgac 261
      ||||  ||||  | | |||  |||  ||  ||||  | |  |  |  |  |  |
Db      710  CCCCTCCCCACCCCGCCCCACCCCCCCCCCCCCACCCCCCCCCCACCCCCCCCCCAC 769

Qy      262  ctctcc 267
      | | |
Db      770  CCCCC 775

```

RESULT 9
CNS016KT
LOCUS CNS016KT 1013 bp DNA GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC BACN16J16 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL106871

VERSION AL106871.1 GI:5624218
 KEYWORDS GSS.
 SOURCE fruit fly.
 ORGANISM Plasmid *Drosophila melanogaster*
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; *Drosophila*.
 REFERENCE 1 (bases 1 to 1013)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 COMMENT Determination of this BAC-end sequence was carried out as part of a
 collaboration with the European *Drosophila* Genome Project (EDGP) -
 http://www.edgp.ebi.ac.uk -. This *Drosophila melanogaster* BAC
 library (Dros BAC) was made by Alain Billaud at CEPH (Centre
 d'Etude du Polymorphisme Humain) with funding provided by a MRC
 project grant. The DNA was prepared from embryos by Alain Bucheton
 and Genevieve Payan. It has been constructed in the vector
 pBeloBAC11.
 FEATURES Location/Qualifiers
 source 1..1013
 /organism="Drosophila melanogaster"
 /plasmid="pBeloBAC11"
 /db_xref="taxon:7227"
 /clone_lib="DrosBAC"
 /clone="BACN16J16"
 /note="end : SP6"
 BASE COUNT 132 a 191 c 148 g 131 t 411 others
 ORIGIN

Query Match 10.6%; Score 43.6; DB 13; Length 1013;
 Best Local Similarity 16.3%; Pred. No. 0.67;
 Matches 53; Conservative 111; Mismatches 162; Indels 0; Gaps 0;

Qy 22 cttctctgctcaatcaattacacaacaagagcattctagatttgagttcatcctagcgat 81
 | | ||| ::: : |:: | : ::| : : : : : : : : : :
 Db 221 CNMCCCTGTNMMMMCMCMCCMSMCCMTMMSCCNMNCMMYSCSCCCCMGMMMMCMNMM 280
 Qy 82 accaatacacccatcccaacactccaaaccaacacattcaaccaaaccaccacaaca 141
 :| : : : : : : : | : : | : : : : : : : : : : : : : :
 Db 281 CMCCMNNMMMSMMMVMNMNCGMGMTAMCMTGNNTNGMMGTTHMMMGNTTTMNGMGGMN 340
 Qy 142 atgccttcagtaaccagggcccggtctcatgtggcgtagcgtagcgccgcggcatggttagag 201
 | :
 Db 341 GNGMNNNNCNGMNNNNMMMMMMMMMGCMGGGGMGMGMNTCTTGMMGMCNMGMCMNMMNC 400
 Qy 202 cccaccccttcgctcgcaatcccatcaccatgaccctcacgcctggcgcgccgac 261
 :| :
 Db 401 MCMNN 460
 Qy 262 ctctccaagaaagtcgtgaagacaagcactgtcttcttcccttctatgcaggtatcctt 321
 | | : : : | : | : | : | : | : | : | : | : | : | : | : | :
 Db 461 NTGCMNMGMGNGNGGNGGNGGNNMGGTCTYCKTTYCTYCCCGGCGSCKTTTCTSTKCTK 520

Qy 322 ggatggccagtcgcagccgcctggtg 347
 :: || | |::| : : : ||
 Db 521 KKCTGTCTGGKSGSCCBGSGCSTTG 546

RESULT 10

CNS0073W/c

LOCUS CNS0073W 922 bp DNA GSS 03-JUN-1999
 DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #
 BACR14D09 of RPCI-98 library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.

ACCESSION AL066784

VERSION AL066784.1 GI:4945247

KEYWORDS GSS.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 922)

AUTHORS Genoscope.

TITLE Direct Submission

JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)

COMMENT Determination of this BAC-end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila
 melanogaster genome using these BACs. For further information
 please see <http://www.fruitfly.org> The BDGP Drosophila
 melanogaster BAC library was prepared by Kazutoyo Osoegawa and
 Aaron Mammoser in Pieter de Jong's laboratory in the Department of
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
 NY. The library is named RPCI-98 and was constructed by partial
 EcoRI digestion of Drosophila DNA provided by the BDGP from the
 isogenic strain y2; cn bw sp, the same strain used for the BDGP's
 P1 and EST libraries. A more detailed description of the library
 and how to order individual BAC clones, the entire library, or
 filters for hybridization from the BACPAC Resource Center can be
 found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source

Location/Qualifiers

1. .922

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone_lib="RPCI-98"

/clone="BACR14D09"

/note="end : TET3"

BASE COUNT 223 a 95 c 109 g 221 t 274 others

ORIGIN

Query Match 10.4%; Score 42.8; DB 13; Length 922;
 Best Local Similarity 22.2%; Pred. No. 1;
 Matches 65; Conservative 105; Mismatches 119; Indels 4; Gaps 1;

Qy 1 agcaaaagcatagagatccatcttctctgctcaatcaattacacaacaagagcattctag 60
 | ::|| :: : : | ::| | : | : ::| :: : ::|| | :: :


```

Db      855 AMNMMMACMMMMCMMACMMAMCCMMACMMMAMAMMMMMMMAMMAMCACMAMMMACACMC 796
Qy      61 atttgagttcatcctagcgataccaatacacccatcccaacactccaaaccaaccaacac 120
      |      :: :: : | : :||| :||| |: :||::|:: ::| :|::| :||:::
Db      795 AMMMCMMMMMMMMMCMCMCMCCACMMMACACMAMCCMMCMCMACMCMMAAMAAMMM 736
Qy     121 ttcaaccaaaccaccacaacaatgccttcagtaaccc---aggcccgctctcatgtggcg 176
      ||:: :||:::|:|::| : : : :||: | : : | : : :
Db      735 MACAMMAMAAMMMMMAMAAMMAAMMAAMMMAMMCCMCCMAAMAMAMMACMMCMCAMMM 676
Qy     177 tagcgctcgcccgcggcgatggtagagccccacccttcgctcgcaatcccatcaccatgac 236
      : | : : : : : : | | :|:|::| : : : :| :| :| :| :| :| :|
Db      675 MMMCAMMMMMAMMAMMMCCCAACAMMCMCACMMCMCAAMAMMMMACMMMAMMAAMMACMC 616
Qy     237 ccctcacgcctggcgcgcgccgcgcacctctccaagaaagtcgtgaagacaagca 289
      : ||| | :| | :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      615 AAMMCACAAAMMACMCMCMCASACACAMVMRAAMMMCCCASAMAAMMAMMA 563

```

RESULT 11

CNS016LW

LOCUS CNS016LW 1101 bp DNA GSS 26-JUL-1999

DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC BACN16J16 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL106910

VERSION AL106910.1 GI:5624430

KEYWORDS GSS.

SOURCE fruit fly.

ORGANISM Plasmid Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1101)

AUTHORS Genoscope.

TITLE Direct Submission

JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - <http://www.edgp.ebi.ac.uk> -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.

FEATURES

source

Location/Qualifiers

1. .1101

/organism="Drosophila melanogaster"

/plasmid="pBeloBAC11"

/db_xref="taxon:7227"

/clone_lib="DrosBAC"

/clone="BACN16J16"

/note="end : T7"

BASE COUNT 222 a 80 c 146 g 113 t 540 others

ORIGIN

Query Match 10.2%; Score 42; DB 13; Length 1101;
 Best Local Similarity 11.8%; Pred. No. 1.7;
 Matches 41; Conservative 128; Mismatches 173; Indels 4; Gaps 1;

```

Qy      34 atcaattacacaacaagagcattctagatttgagttcatcctagcgataccaatacaccc 93
      :      :      :      :      :      :      :      :      :      :
Db      517 MNNNNNNNNNNNNNNNNMMGGMGNNNNNNNNMCNMMTNNNNNNNNNNNNNNNNNTNNNTNMNTT 576

Qy      94 atcccaacactccaaaccaaccaacacttcaaccaaaccaccacaacaatgccttcagta 153
      :      :      :      :      :      :      :      :      :      :
Db      577 NNNNNNNNNMCMNNNNNNNNMCMNNNNNNNNMCMNNNNNNNNMCMNNNNNNNNMCMNNNNNNNNM 636

Qy     154 acccaggcccgtctcatgtggcgtagcgtgcgccgcggcatggtagagccccacccttc 213
      :      :      :      :      :      :      :      :      :      :
Db      637 MNNNNNNNNNNMCMNNMGNCNNNNNNNNCNCNNMMNCNCNNMMCMNNMCMNNNNNNNNMCMN 696

Qy     214 gctcgcaatcccatcaccatgaccctcacgcctggcgcgccgcgcacctctccaagaaa 273
      :      :      :      :      :      :      :      :      :      :
Db      697 NNNNNNNNNMCMNNMCMCMGCMHGMKMTGTMTKTTTMMWTTGHMGMHHGMMMNTVMMGGWMT 756

Qy     274 gtcgtgaagacaagcactgtcttcttcccttctatgcaggtatccttgatggccagtc 333
      ||: :      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db      757 GTMMKMKMGHMTG----WGTGKTMGHMMMNMDTMTKMVMKTTTMMAGMMRNVGMGAKGT 812

Qy     334 gcagccgcctggtggttcaacggaaacatgtgactcttccaaatgg 379
      :||: :      :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db      813 MTWGVMMVRKMMVDMGTRGTGHHGMMKWTVTGMGTAKKGTNGMTGD 858
  
```

RESULT 12

CNS011EQ
 LOCUS CNS011EQ 748 bp DNA GSS 26-JUL-1999
 DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC
 BACN06J14 of DrosBAC library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.
 ACCESSION AL100172
 VERSION AL100172.1 GI:5611783
 KEYWORDS GSS.
 SOURCE fruit fly.
 ORGANISM Plasmid Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 748)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 COMMENT Determination of this BAC-end sequence was carried out as part of a
 collaboration with the European Drosophila Genome Project (EDGP) -
 http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
 library (Dros BAC) was made by Alain Billaud at CEPH (Centre
 d'Etude du Polymorphisme Humain) with funding provided by a MRC
 project grant. The DNA was prepared from embryos by Alain Bucheton

and Genevieve Payan. It has been constructed in the vector pBeloBAC11.

```
FEATURES             Location/Qualifiers
    source             1. .748
                        /organism="Drosophila melanogaster"
                        /plasmid="pBeloBAC11"
                        /db_xref="taxon:7227"
                        /clone_lib="DrosBAC"
                        /clone="BACN06J14"
                        /note="end : SP6"
BASE COUNT            174 a      349 c      35 g      62 t      128 others
ORIGIN
```

Query Match 10.2%; Score 41.8; DB 13; Length 748;
Best Local Similarity 44.1%; Pred. No. 1.8;
Matches 100; Conservative 15; Mismatches 112; Indels 0; Gaps 0;

```
Qy      82 accaatacacccatcccaacactccaaaccaacacattcaaccaaaccaccacaaca 141
      ||:| : |||| ||| | | || ||| | | |||| ||| ||| |
Db      134 ACSACCCMSCCCACACCACCCSCCCSCACCCCAACSACCCCAAACCCCCCAACACCCCC 193

Qy      142 atgccttcagtaacccaggcccggtctcatgtggcgtagcgctcgcccgcgccatggttagag 201
      || || |:|| |:|: || | | ||| | || |
Db      194 CACCCCSCAASCASSCAACSSSCACCAACCCCCCCCCCAACCCCCCCCCACCCCCCAA 253

Qy      202 cccaccccttcgctcgcaatcccatcaccatgacccctcacgcctggcgcgccgac 261
      ||||| ||| : | || ||| |||: || | |:| : |:|:| |
Db      254 CCCCACCCCAAMCCCAACCCCAACCCCAACSCCCACAACCASCSCCACAACSCSCSCCCCC 313

Qy      262 ctctccaagaaagtcgtgaagacaagcactgtcttcttcccttcta 308
      | | | |: | | ||| || | | |:| | |
Db      314 ACCSCCCACSCASCACCCCAACCACCCACCCCCCSCCCCCAA 360
```

```
RESULT 13
CNS01FKL
LOCUS      CNS01FKL      1028 bp      DNA      GSS      01-JUN-2001
DEFINITION Anopheles gambiae GSS T7 end of clone 04L01 of NotreDamel library
            from strain PEST of Anopheles gambiae (African malaria mosquito),
            genomic survey sequence.
ACCESSION  AL141958
VERSION    AL141958.1 GI:7000076
KEYWORDS   GSS.
SOURCE     African malaria mosquito.
ORGANISM   Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
            Culicoidea; Anopheles.
REFERENCE  1 (bases 1 to 1028)
AUTHORS    Genoscope.
TITLE      Direct Submission
JOURNAL    Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
REFERENCE  2 (bases 1 to 1028)
AUTHORS    Roth,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.
```


ORGANISM Chlamydomonas reinhardtii
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 Chlamydomonadaceae; Chlamydomonas.

REFERENCE 1 (bases 1 to 1737)

AUTHORS Grossman,A., Davies,J., Federspiel,N., Harris,E., Hauser,C.,
 Lefebvre,P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D.

TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model,
 Unicellular System for Analyzing Gene Function and Regulation in
 Vascular Plants; project phase 3

JOURNAL Unpublished (2000)

COMMENT Contact: Charles Hauser
 DCMB Box 91000
 Duke University
 Durham, NC 27708-1000
 Tel: 919 613 8159
 Fax: 919 613 8177
 Email: chauser@duke.edu.

FEATURES Location/Qualifiers

source 1. .1737
 /organism="Chlamydomonas reinhardtii"
 /strain="CC-1690 wild type mt+ 21gr"
 /db_xref="taxon:3055"
 /clone_lib="C. reinhardtii CC-1690, Stress condition I,
 normalized, Lambda Zap II"
 /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
 XhoI; This library, constructed by John Davies and Jeffrey
 McDermott, combines cDNAs from CC-1690 cells grown to
 mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min,
 1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr
 , 4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was
 purified from each sample, pooled and cDNA synthesized.
 The cDNA was directionally cloned into lambda Zap II
 (Stratagene) in the EcoRI (5') and XhoRI (3') sites.
 pBluescript II SK- plasmids were excised from the lambda
 ZAP clones by superinfection with ExAssist (Stratagene)
 phage. The library was normalized using method 4 described
 in Bonaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT 506 a 968 c 129 g 117 t 17 others

ORIGIN

Query Match 10.1%; Score 41.4; DB 11; Length 1737;
 Best Local Similarity 53.8%; Pred. No. 2.8;
 Matches 84; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

Qy 88 acacccatcccaacactccaaaccaaccaactttcaaccaaaccaccacaacaatgcct 147
 ||||| || |||| | |||| | |||| | ||||| ||||| |

Db 609 ACACCCACCCACACACACACCACCACCACACACACACCCACCACCACCACAACCAAACAA 668

Qy 148 tcagtaacccaggcccggtctcatgtggcgtagcgtagcgccgcggcatggtagagccccac 207
 | |||| | |||| | || || || |||| |||| |||| | ||||

Db 669 ACTCAAACCAACAACCGACCCACACACCGCTCAGCCCCCGTCACATCCTCGTCGCACAC 728

Qy 208 cccttcgctcgcaatcccatcaccatgaccctcac 243
 | | |||| | || |||| | || |

Db 729 ACACACTCTGCCAGACCCACCCANCAACCAACCC 764

Search completed: February 7, 2002, 08:20:58
Job time: 18135 sec